

Page 5

Db	746	TGATAGTCAAAATGATCATAAAGCCAGGTTTGCTTCACACTTCCTCGAAATTTTACTCAC	805
Qy	2832	agatcatttggcaacaagcatagcttacttatgttttagggactgaacaattttatttgggaa	2891
Db	806	AGATCATTTGGCAACAAGCATAGCTTACTTATGTGTTAGGGACTGAACAATTTATTGGAA	865
Qy	2892	gcaaaccttttatatgctcagaaagtacattttaaaagatgactacttacgcaggaggatgc	2951
Db	866	GCAAACTCTTTATATGCTAGAAAGTACATTTAAAAGATGACTACTTACGCAGGAGATGC	925
Qy	2952	aggtctctctaaacgcataatgtatgtatgttaggcacatgtagtgagtgcataatg	3011
Db	926	AGGTCTCTCTAAACGCATGAATGTATGTAGTGTGTAGGCACGTACTGAGTGATATATG	985
Qy	3012	ctccacactacgtcttgataaaacacaaacctcagttatctcagttattaggcacactagttt	3071
Db	986	CTCCACACTAGCTGTGATATAACACAAAGCTCAGTATTCAGTTATTAGGCACACTAGTTTT	1045
Qy	3072	atacgcaactactgcttcacatagtagactgtttgttgccaataaactcttgaattgtct	3131
Db	1046	ATAGCGCAACTACTGCTTACATAGTAGACTGTTTGTGTGCCAATAACTTTGAATGTTCT	1105
Qy	3132	ttaaagaaactgaggttcagatacacataccatggaaaaatcttactttcttctgttact	3191
Db	1106	TTAAAGAAAACTGAGGTTTCAGATACACATACCATGGAAAACTTTACTTTCTTGTACT	1165
Qy	3192	acacaaagctattttaagaagatgctatgttggagaaggcggaagtgtactatatga	3251
Db	1166	ACACAAAGCTATTTTAAAGAAAGATGCTATGTTGGGAGAAAGGCGGAAGTGTACTATATGA	1225
Qy	3252	cataatcaat	3261
Db	1226	CATAATCAAT	1235

3

RESULT
HSIMPGL3
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SEGMENT
SOURCE

HSIMPGL3 816 bp DNA
Homo sapiens interphotoreceptor matrix gene (IPM150), exon 13.
AF011772
AF011772.1 GI:3800727
13 of 17
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 816)
Felbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
Schmid, M. and Weber, B.H.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE

Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate
for 6q-linked retinopathies
Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
98358139
2 (bases 1 to 816)
Gehrig, A., Felbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussels, I.E.
and Weber, B.H.F.
Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRL)
Unpublished
3 (bases 1 to 816)
Felbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F. .
Direct Submission
Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
Location/Qualifiers
1. .816
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"

FEATURES
source

FEATURES
source
1. .4368
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
190. .3909
/gene="Pg10.2"
190. .3909
/gene="Pg10.2"
/note="expressed in the pineal gland and retina"
/codon_start=1
/db_xref="GI:166797"
/translation="MPLPLRSLGLLILFLGGNLVSVSEIQDRMHAVVLSPKRS
TDLSPTRKRLQDATERRRRLRRRLILFNGVRICPSTVAEAVANHYKFKAR
VQCEIWEAFRTFDRLPGLREYQYWMNLCEDGVTSVFEMGTQFSQSVHRHLEKIL
TYKAESSCKDQACQGLSPVPIGTSTLAGAVSSAPGAASERAAPOESIL
NEIENVTQPTPAEQIAEFSIQLLGKQYSELRDPSALYLLVEERISEVERAFT
GLPGYKGIHVLDFRSPKNGSIDHYAVTFNGEALSNWTWDLISLHKNKVENHGLVE
LDKPTAVTISNFRDIYIAETLHQPLMNSLNDPKSLQNLINVRGLLPQTEIIV
NTQSSLVQVTSILNDTLQAEWLGADESITTTTTISPPGSSGPPSATGRELHSE
STLGDIVSTPKLASPKVVLSPSSLIQVPRKTPVPMEDSDMLTSSPHLTSSVIEDLA
ILEDNTSESDVSIDLVPSSLIQVPRKTPVPMEDSDMLTSSPHLTSSVIEDLA
KDITTPSGLDISASVSDKLDVSPHPTDTSVEKELFESGLSGSKNDVVDWPKSE
TSLEKTPLSKSWSEEDTLPTESIEKLMYFTEQMIPEPSAHRIQDGIPTTEES
HYRTPIFAESATQPTSLISKHTSDVPDIDSVYTRKAPFLIANTASTKETDEV
NTLLKGVQTEPSPKGLDSKISVARDPMQPVWTLFESDVTWARTSLGLSRDITL
VSTPESADRLWKAQSTQPAELPSTHSLQLEDEIVMAQNIISLELDQVGTQDYQPEL
TOEQNGVDVSVEMPTHVYTEMPLVAOPTKGGVLSRTQAGALVVEFSRLVTNMLFS
EDLFNKSLEYKALEPQRELELVYQNSLQSFQNLILNFRNGSIVNRSVKFAESV
PPVNNALYMLEIDCTAYQTMNDIDKYSLDVSGDDANPKFOACNEFCECLVNP
WSGAEKCHPGILSDPELPCQVCDLQDFCLNGDKDQVMPGHAICRCEVGSNNWY
RGHCEVFPFVIGITATVSVLLVASVFFLAKMLQANVRERQRPTNRPQD
SLSSVENAKNYPVIESLAGOEYKPYSPHPFYSASEEYIGLSREIIRQMYESS
DLSKEIEQMRILELYANDPFAAFVREHEMEL"

TITLE
JOURNAL
COMMENT

1228 a 986 c 1024 g 1130 t

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 0.9%; Score 38; DB 12; Length 4368;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3222 gccacaccttcgaagtgcctgaatgaatttc 3259
Db 3223 gccacaccttcgaagtgcctgaatgaatttc 3260

AC026980 65674 bp DNA HTG 25-MAR-2000
Homo sapiens chromosome 21 clone RP11-65K18 map 21, LOW-PASS
SEQUENCE SAMPLING.

AC026980
AC026980.1 GI:7328857
HTG; HTGS_PHASE0.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 65674)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 21, clone RP11-65K18
Unpublished

2 (bases 1 to 65674)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,J., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Menes,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L8662
Center clone name: 65_K_18

* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 664: contig of 664 bp in length
665 764: gap of 100 bp
765 1456: contig of 692 bp in length
1457 1556: gap of 100 bp
1557 2235: contig of 679 bp in length
2236 2335: gap of 100 bp
2336 3002: contig of 667 bp in length
3003 3102: gap of 100 bp
3103 3789: contig of 687 bp in length
3790 3889: gap of 100 bp
3890 4593: contig of 704 bp in length
4594 4693: gap of 100 bp
4694 5378: contig of 685 bp in length
5379 5478: gap of 100 bp
5479 6167: contig of 689 bp in length
6168 6267: gap of 100 bp
6268 6942: contig of 675 bp in length
6943 7042: gap of 100 bp
7043 7711: contig of 669 bp in length
7712 7811: gap of 100 bp
7812 8488: contig of 677 bp in length
8489 8588: gap of 100 bp
8589 9285: contig of 697 bp in length
9286 9385: gap of 100 bp
9386 10069: contig of 684 bp in length
10070 10169: gap of 100 bp
10170 10845: contig of 676 bp in length
10846 10945: gap of 100 bp
10946 11632: contig of 687 bp in length
11633 11732: gap of 100 bp
11733 12421: contig of 689 bp in length
12422 12521: gap of 100 bp
12522 13206: contig of 685 bp in length
13207 13306: gap of 100 bp
13307 14002: contig of 696 bp in length


```

Query Match      8.2%; Score 266; DB 91; Length 561;
est Local Similarity 99.3%; Pred. No. 2.4e-132;
Matches 416; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1226 caatcttggatgtggggacaattccagttccactgatgaaattgctggatcactgccagcc 1285
|||||
465 CAATCTTTGGATGTGGGGACAATTCAAGTTCAGTATGAAATCTGCTGGATCAGTCGACGC 406
|||||

1286 ttgggctctgcaccccaatcagagctgccacatcttttgcgttataacagaggatgct 1345

```

Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavet@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M3 Reverse.
 Location/Qualifiers
 1. .298
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):194417"


```
RESULT 2
AA721009/c 435 bp mRNA EST 22-JAN-1998
LOCUS nx89e03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269436 3
DEFINITION similar to TR:P70628 P70628 PGL0.2. ; mRNA sequence.
ACCESSION AA721009.1 GI:2737144
VERSION AA721009.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 435)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2151557.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1678 Std Error: 0.00
Seq Primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 416.
Location/Qualifiers
1. 435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1269436"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAACTGAAGTGGGAGCGCGCTCATTTTTCATTTTTCATTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
128 a 109 c 74 g 124 t
BASE COUNT
ORIGIN
Query Match 4.9%; Score 204; DB 30; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.6e-97;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3223 ccaaccctgcagttccagctgctgaattttcagagtgctcgtggaacccctgga 3282
|||||
Db 425 CCACCCCTGCAAGTTTCAAGGCTGTAAATTTTCAGAGTCTCTGGTCAACCCCTGGA 366
|||||

QY 3283 gtggagaagcaatgcagatgcttcctcctggataccctgagtggtggaacgcccctgtc 3342.
|||||
Db 365 GTGGAGAGCAAGATGCAGATGCTTCCCTGGATACCTGAGTGTGGAGAACGCCCTGTC 306
|||||

QY 3343 agagtcctctgacactacagcctgactctgttgatgatgaaagtgtgacattatgc 3402
|||||
```

```
Db 305 AGAGTCTGTGACCTACAGCCTGACTTCTGCTGAATGATGAAAGTGTGACATTATGC 246
QY 3403 ctgggcacggggccattgttaggt 3426
|||||
Db 245 CTGGGCACGGGGCCATTTGTAGGT 222
|||||

RESULT 3
AA744481/c 435 bp mRNA EST 22-JAN-1998
LOCUS ny25d01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1272769 3'
DEFINITION similar to TR:P70628 P70628 PGL0.2. ; mRNA sequence.
ACCESSION AA744481
VERSION AA744481.1 GI:2783245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 435)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2045613.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1660 Std Error: 0.00
Seq Primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 378.
Location/Qualifiers
1. 435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1272769"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAACTGAAGTGGGAGCGCGCTCATTTTTCATTTTTCATTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
127 a 108 c 74 g 126 t
BASE COUNT
ORIGIN
Query Match 4.7%; Score 194; DB 30; Length 435;
Best Local Similarity 100.0%; Pred. No. 5.6e-92;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3233 caagtttcaggccctgtaataatttcagagtgctcgtggaacccctggaagc 3292
|||||
Db 415 CAAAGTTTCAGGCCGTGAATGAATTTTCAGAGTGTCTGGTCAACCCCTGGAGTGAAGC 356
|||||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 17:40:05 ; Search time 2736.15 seconds
(without alignments)
6712.693 Million cell updates/sec

Title: US-09-183-972-5
Perfect score: 4165
Sequence: 1 cgggywaytttgaaggagca.....aaactacgttaaaaaaaa 4165

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 5247842 seqs, 2204914090 residues
Word size : 0
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : EST:*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*
44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	gb_est33:*
60:	gb_est34:*
61:	gb_est35:*
62:	gb_est36:*
63:	gb_est37:*
64:	gb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	gb_est39:*
70:	gb_est40:*
71:	gb_est41:*
72:	gb_est42:*
73:	gb_est43:*
74:	gb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	gb_est45:*
80:	gb_est46:*
81:	gb_est47:*
82:	em_est35:*
83:	em_est36:*
84:	em_est37:*
85:	gb_est48:*
86:	gb_est49:*
87:	gb_est50:*
88:	gb_est51:*
89:	gb_est52:*
90:	gb_est53:*
91:	gb_est54:*
92:	gb_est55:*
93:	gb_gss1:*
94:	gb_gss2:*
95:	gb_gss3:*
96:	gb_gss4:*
97:	em_gss1:*
98:	em_gss2:*
99:	em_gss3:*
100:	em_gss4:*
101:	gb_gss5:*
102:	gb_gss6:*
103:	gb_gss7:*
104:	gb_gss8:*
105:	gb_gss9:*
106:	em_gss5:*
107:	em_gss6:*
108:	em_gss7:*
109:	em_gss8:*
110:	em_gss9:*
111:	em_gss10:*
112:	em_gss11:*
113:	gb_gss10:*
114:	gb_gss11:*
115:	em_gss12:*
116:	gb_gss12:*

117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	214	5.1	435	31	AA815118	AA815118 oa88h10.s
C 2	204	4.9	435	30	AA721009	AA721009 nx89e03.s
C 3	194	4.7	435	30	AA744481	AA744481 ny25d01.s
C 4	162	3.9	383	30	AA736980	AA736980 nx89a04.s
C 5	120	2.9	202	30	AA766994	AA766994 oa41c01.s
C 6	113	2.7	534	104	AQ560890	AQ560890 HS_2089_B
C 7	62	1.5	447	43	AI739559	AI739559 w135a06.x
C 8	23	0.6	353	38	AI346256	AI346256 qp49f04.x
C 9	22	0.5	204	80	CI5921	CI5921 CJ5921 Clon
C 10	22	0.5	221	37	AI248036	AI248036 qh63f09.x
C 11	22	0.5	228	21	AA112176	AA112176 zm64h03.s
C 12	22	0.5	249	21	AA084038	AA084038 zm64h03.r
C 13	22	0.5	253	25	AA365447	AA365447 EST76221
C 14	22	0.5	266	32	AA876973	AA876973 ny49c05.s
C 15	22	0.5	271	32	AA860355	AA860355 aj59b04.s
C 16	22	0.5	273	34	AI033165	AI033165 ow94c06.s
C 17	22	0.5	280	29	AA629627	AA629627 ac21h02.s
C 18	22	0.5	282	38	AI381763	AI381763 te40h02.x
C 19	22	0.5	287	30	AA774219	AA774219 ab55f03.s
C 20	22	0.5	292	89	T30212	T30212 EST12890 Hu
C 21	22	0.5	292	89	T51656	T51656 vb28h12.s1
C 22	22	0.5	300	36	AI202486	AI202486 qs69g05.x
C 23	22	0.5	305	44	AI830060	AI830060 wj58g10.x
C 24	22	0.5	312	36	AI193938	AI193938 qe73d08.x
C 25	22	0.5	313	64	AW173122	AW173122 xj83e11.x
C 26	22	0.5	321	89	T33508	T33508 EST58101 Hu
C 27	22	0.5	322	46	AI936694	AI936694 wp68f05.x
C 28	22	0.5	324	21	AA096111	AA096111 l8239.seq
C 29	22	0.5	324	43	AI699436	AI699436 wa70c10.x
C 30	22	0.5	326	91	W48841	W48841 zc42h08.s1
C 31	22	0.5	329	74	AW575664	AW575664 UI-HF-BM0
C 32	22	0.5	330	28	AA587626	AA587626 nm95c08.s
C 33	22	0.5	333	38	AI372439	AI372439 EST175269
C 34	22	0.5	334	31	AA777008	AA777008 zf28a06.s
C 35	22	0.5	340	81	D52481	D52481 HUM079D09B
C 36	22	0.5	341	31	AA846109	AA846109 ak83g02.s
C 37	22	0.5	341	33	AA961248	AA961248 om66b12.s
C 38	22	0.5	342	70	AW316570	AW316570 wx08h10.x
C 39	22	0.5	344	37	AI275925	AI275925 qw05b02.x
C 40	22	0.5	344	38	AI359644	AI359644 qy3c05.x
C 41	22	0.5	345	43	AI698523	AI698523 wa79a03.x
C 42	22	0.5	349	38	AI372437	AI372437 EST175267
C 43	22	0.5	349	89	T30585	T30585 EST19020 Hu
C 44	22	0.5	354	30	AA772838	AA772838 ae74g05.s
C 45	22	0.5	355	33	AA973982	AA973982 oq05g11.s

ALIGNMENTS

RESULT 1
AA815118/c 435 bp mRNA EST
LOCUS oa88h10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1319395 3
DEFINITION similar to TR:P70628 P70628 PG10.2.; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA815118
AA815118.1 GI:2884714
EST.
human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 435)

NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Jan 17, 1998 this sequence version replaced gi:2045394.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbp/image/image.html

Insert Length: 1672 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 381.

Location/Qualifiers

1..435

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1319395"

/clone_lib="NCI_CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD+),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'

] Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 127 a 109 c 74 g 125 t

ORIGIN

Query Match 5.1%; Score 214; DB 31; Length 435;

Best Local Similarity 100.0%; Pred. No. 1.2e-102;

Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3213 ggtgatgaagcacaaccccttcgaagtttcaggctgttaatttcagagtgctgttc 3272

Db 435 GGTGATGAAGCAACCCCTTCGAAGTTTCAGGCTGTAAATTTTCAGAGTGTCTGTC 376

QY 3273 aaccccttgagtgagagcaaaagtcagatgtcccttgatcaccttgagtggaagaa 3332

Db 375 AACCCCTGGAGTGAGGAAGCAAGTGCAGATGCTCCCTGGATACCTGAGTGTGGAAGAA 316

QY 3333 cggccctgtcagagttctgtgacctacagcctgacctctgtctgctgaatgaggaagtg 3392

Db 315 CGGCCCTGTGAGAGTCTCTGTGACCTACAGCTGACCTTCGTCGTGATGATGGAAGTGT 256

QY 3393 gacattatgctgggcacggggccatttgtaggt 3426

Db 255 GACATTATGCTGGGCACGGGGCCATTGTAGGT 222

```
RESULT 2
AA721009/c 435 bp mRNA EST 22-JAN-1998
LOCUS
DEFINITION
nX89e03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269436
similar to TR:P70628 P70628 PG10.2.; mRNA sequence.
AA721009
ACCESSION
AA721009
VERSION
AA721009.1 GI:2737144
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 435)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 19, 1998 this sequence version replaced gi:2151557.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 1678 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 416.
Location/Qualifiers
1. .435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1269436"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-GTGTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 128 a 109 c 74 g 124 t
ORIGIN
Query Match 4.9%; Score 204; DB 30; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.6e-97;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3223 ccaacccttcgaagtttcaggcctgaatgaatttcagagtgctcgtgtaaccccttga 3282
|||||
DB 425 CCAACCTTCAGTTTCAGGCTGTAATGAATTTTCAGAGTGTCTGTTCAACCCCTGGA 366
QY 3283 gtggagaacaaagtacagatgttccctggatacctcagtggtgggaagacggccctgtc 3342.
|||||
DB 365 GTGGAGAAGCAAGTGCAGATGTTCCCTGGATACCTGTAGTGTGGAGAAGACGGCCCTGTC 306
QY 3343 agagttctgtacacctacagcctcctcgttgatgatgaagtgacattatgc 3402
|||||

Db 305 AGAGTCTCTGTGACCTACAGCCTGACTTCTGCTTGAATGATGGAAGAGTGTGACATTATGC 246
QY 3403 ctgggcacggggccattttaggt 3426
|||||
Db 245 CTGGCAGCGGGCCATTGTAGGT 222

RESULT 3
AA744481/c 435 bp mRNA EST 22-JAN-1998
LOCUS
DEFINITION
ny25d01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1272769 3'
similar to TR:P70628 P70628 PG10.2.; mRNA sequence.
AA744481
ACCESSION
AA744481
VERSION
AA744481.1 GI:2783245
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 435)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 19, 1998 this sequence version replaced gi:2045613.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 1660 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 378.
Location/Qualifiers
1. .435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1272769"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 127 a 108 c 74 g 126 t
ORIGIN
Query Match 4.7%; Score 194; DB 30; Length 435;
Best Local Similarity 100.0%; Pred. No. 5.6e-92;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3233 caatttcaggccctgaatgaatttcagagtgctcgtgtaaccccttgagtggaagc 3292
|||||
Db 415 CAAGTTTCAGGCCCTGTAATGAATTTTCAGAGTGTCTGTTCAACCCCTGGAGTGAAGC 356
```

Qy 3293 aaagtgcagatgctccctgatacctgagtggtggaagcgcctgctcagatgctctg 3352
 |||
 Db 355 AAGTGCAGAGCTTCCTCGATACCTGAGTGTGGAAGAACGCCCTGFCAGAGTCTCTG 296
 |||
 Qy 3353 tgacctacagcctgactctctctgctgaatgatggaagtgtgacattatgctctggcaggg 3412
 |||
 Db 295 TGACCTACAGCCTGACTCTCTGCTGAATGATGGAAGTGTGACATTATGCTTGGGCACGG 236
 |||
 Qy 3413 ggcatttgtaggt 3426
 |||
 Db 235 GGCCATTGTAGGT 222

RESULT 4
 AA736980/c 383 bp mRNA EST 22-JAN-1998
 LOCUS nX89a04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269390 3',
 DEFINITION mRNA sequence.
 ACCESSION AA736980
 VERSION AA736980.1 GI:2767255
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 383)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1623 Std Error: 0.00
 Seq primer: -40ml3 fwd. Et from Amersham
 High quality sequence stop: 347.
 Location/Qualifiers
 1..383
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1269390"
 /clone_lib="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p7T73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGGCGGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p7T73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES

source
 112 a 96 c 64 g 111 t
 BASE COUNT
 ORIGIN
 Query Match 3.9%; Score 162; DB 30; Length 383;
 Best Local Similarity 100.0%; Pred. No. 6.8e-75;
 Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3265 gtctgtcaacccctggagtgagaaagcagatgctccctgatacctgagtg 3324
 |||
 Db 383 GTCGTGTCACCCCTGGAGTGGAGAGCAAAAGTCAGATGCTTCCCTGGATACCTGAGTG 324
 |||
 Qy 3325 tggaagaacggccctgtcagagtcctctgtgacctacagcctgactctctggaatgatg 3384
 |||
 Db 323 TGAAGAACGGCCCTGTGAGAGTCTCTGTGACCTACAGCCTGACTTCTGCTTGAATGATG 264
 |||
 Qy 3385 gaaagtgtgacattatgctctgggcaacggggccattttaggt 3426
 |||
 Db 263 GAAAGTGTGACATTATGCTTGGGCACGGGGCCATTGTAGGT 222

RESULT 5
 AA766994/c 202 bp mRNA EST 08-FEB-1998
 LOCUS oa41c01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1307520 3',
 DEFINITION similar to TR:P70628 P70628 PG10.2.; mRNA sequence.
 ACCESSION AA766994
 VERSION AA766994.1 GI:2819575
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 202)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
 Insert Length: 1669 Std Error: 0.00
 Seq primer: -40mi3 fwd. Et from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers
 1..202

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1307520"
 /clone_lib="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p7T73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGGCGGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p7T73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

50 a 60 c 46 g

ORIGIN

Query Match 2.9%; Score 120; DB 30; Length 202;
 Best Local Similarity 99.4%; Pred. No. 1.8e-52;
 Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3236 gtttcagcctgttaatttcagagtggtgtgtcaaccctcggagtggaagcaaa 3295
 |||||
 Db 179 GTTTCAGCCCTGTAATGAAATTTTCAGAGTGTCTGCTCAACCCCTGGAGTGGAGCAAA 120
 |||||

QY 3296 gtgcagatgctccctgcatacctagtggtgaagaacgcccctgtcagagtcctgtga 3355
 |||||
 Db 119 GTGCAGATGCTCCCTCGATACCTCAGTGTGGATGAACCGCCCTGTCAGAGTCTCTGTGA 60
 |||||

QY 3356 cctcagcctgactctctgctgaatgatggaagtgatgacattatcctgg 3406
 |||||
 Db 59 CCTACAGCTGACTTCTGCTTGAATGATGGAAGTGTGACATATATGCTGG 9
 |||||

RESULT 6
 A0560890
 LOCUS
 DEFINITION HS_2089_B2_A08_T7C CIT Approved Human Genomic Sperm Library D
 Homo sapiens genomic clone Plate=2089 Col=16 Row=B, genomic
 survey sequence.
 ACCESSION A0560890

VERSION A0560890.1 GI:4920558
 KEYWORDS GSS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 534)
 MAHAIAS G.G., WALLACE J.C., SMITH K., SWARTZELL S., HOLZMAN T.,
 KELLER A., SHAKER R., FURLONG J., YOUNG J., ZHAO S., ADAMS M.D. and
 HOOD L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 93980589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallaceu.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2089 row: B column: 16
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 534.

FEATURES
 source
 1..534
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2089 Col=16 Row=B"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 131 a 107 c 131 g 159 t 6 others
 ORIGIN

Query Match 2.7%; Score 113; DB 104; Length 534;
 Best Local Similarity 100.0%; Pred. No. 1e-48;
 Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3425 gtgcgggtgggtgagactggtgtaccgagcagcactgtgaggaattgtgtcta 3484
 |||||
 Db 164 GTGCGGGTGGGTGAGACTGTTGATCCGAGCAGCACTGTGAGGAATTTGTCTGA 223
 |||||

QY 3485 gcccgatcataggaatcactattgctccgtgggttgactttcttctatct 3537
 |||||
 Db 224 GCCCGTATCATAGGCATCACTATTGCTCCGTGGTGGACTTCTTCTCATCT 276
 |||||

RESULT 7
 A1739559/c
 LOCUS
 DEFINITION

w13sa06.x1 NCI-CGAP-Col6 Homo sapiens cDNA clone IMAGE:2392210 3'
 similar to TR:P70628 P70628 PG10.2.; mRNA sequence.

ACCESSION A1739559
 VERSION A1739559.1 GI:5101540
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 447)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2151197.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.biol.llnl.gov/bbrp/image/image.html
 Insert length: 1596 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 428.

FEATURES
 Location/Qualifiers
 1..447
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2392210"
 /clone_lib="NCI-CGAP-Col6"
 /tissue_type="Colon tumor, RER+"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP-Col0 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (cloneIDs 1057416-1061255, and 1144584-1145351).
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 108 c 89 g 119 t
 ORIGIN

Query Match 1.5%; Score 62; DB 43; Length 447;
 Best Local Similarity 100.0%; Pred. No. 1.7e-21;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1369 ctgaagatctagtttggaacacccaaagtcttcaggtcttcaggacccgtcatcttc 1428
 |||||
 Db 447 CTGAAGATCTAGTTTGGACACCCAAAGTTCAAGTCTTTCAGGCAACCGCTCATCTATTC 388
 |||||

QY 1429 tg 1430
 ||
 Db 387 TG 386

RESULT 8
 A1346256/c

LOCUS A1346256 353 bp mRNA EST 30-DEC-1998
 DEFINITION qp49f04.x1 NCI-CGAP Co8 Homo sapiens cDNA clone IMAGE:1926367 3' similar to contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION A1346256
 VERSION A1346256.1 GI:4083462
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 353)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3035630.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbp/image/image.html
 Seq primer: -400P from Gibco
 High quality sequence stop: 328.

FEATURES
 Location/Qualifiers
 1..353
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1926367"
 /clone_lib="NCI-CGAP Co8"
 /tissue_type="adenocarcinoma"
 /lab_host="DH108"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 113 a 58 c 67 g 115 t
 ORIGIN

Query Match 0.6%; Score 23; DB 38; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4127 tgcgtgttttaataattgaaaa 4149
 |||||||
 DB 232 TGCTGTGTTTAAATGAAAA 210

RESULT 9
 C15921
 LOCUS 204 bp mRNA EST 30-SEP-1996
 DEFINITION C15921 Clontech human aorta polyA+ mRNA (#5572) Homo sapiens cDNA clone GEN-408F04 5', mRNA sequence.

ACCESSION C15921
 VERSION C15921.1 GI:1570628
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 204)
 AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,

Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y.
 Fujiwara et al (1995)
 Unpublished (1995)
 On Jul 7, 1999 this sequence version replaced gi:5407660.
 Contact: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co.,Ltd
 43-10 kagasuno Kawauchi-cho, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.

FEATURES
 Location/Qualifiers
 1..204
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GEN-408F04"
 /clone_lib="Clontech human aorta polyA+ mRNA (#5572)"

BASE COUNT 54 a 54 c 45 g 45 t
 ORIGIN

Query Match 0.5%; Score 22; DB 80; Length 204;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 tatcaccttttggggaagcag 950
 |||||||
 DB 118 TATCCACCTTTGGGAGACGAG 139

RESULT 10
 A1248036/c
 LOCUS 221 bp mRNA EST 01-DEC-1998
 DEFINITION qh63f09.x1 Soares_fetal_liver_spleen_lnfpls_sl Homo sapiens cDNA clone IMAGE:1849385 3', mRNA sequence.

ACCESSION A1248036
 VERSION A1248036.1 GI:3843433
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 221)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2284999.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 2289 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 195.

FEATURES
 Location/Qualifiers
 1..221
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1849385"
 /clone_lib="Soares_fetal_liver_spleen_lnfpls_sl"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subcloned version of the original Soares fetal liver spleen lnfpls library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAGATTAATTAAGATCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

68 a 38 c 51 g 64 t

Query Match 0.5%; Score 22; DB 37; Length 221;

Best Local Similarity 100.0%; Pred. No. 3.8; Mismatches 0; Indels 0; Gaps 0;

Qy 929 tatccaccttttgggaagcag 950

|||||

Db 199 TATCCACCTTTTGGGAAGCAG 178

RESULT 11
LOCUS

AA112176/c 228 bp mRNA EST 23-DEC-1997

DEFINITION zm64h03.s1 Stratagene fibroblast (#937212) Homo sapiens cDNA clone

IMAGE:530453 3', mRNA sequence.

ACCESSION
VERSION

AA112176.1 GI:1664378

KEYWORDS
SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 228)

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.

TITLE
JOURNAL

Generation and analysis of 280,000 human expressed sequence tags

GENOME
MEDLINE

Genome Res. 6 (9), 807-828 (1996)

COMMENT

Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 339 Std Error: 0.00

Seq primer: -40M13 fwd. from Amersham.

FEATURES
source

1. .228
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:3919964"
/db_xref="taxon:9606"
/clone="IMAGE:530453"
/clone_lib="Stratagene fibroblast (#937212)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. W138 cell line. Average insert size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT
ORIGIN

71 a 40 c 51 g 66 t

Query Match

0.5%; Score 22; DB 21; Length 228;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 tatccaccttttgggaagcag 950

|||||

Db 194 TATCCACCTTTTGGGAAGCAG 173

RESULT 12

AA084038

LOCUS

zm64h03.r1 Stratagene fibroblast (#937212) Homo sapiens cDNA clone

IMAGE:530453 5', mRNA sequence.

ACCESSION

AA084038

VERSION

AA084038.1 GI:1626094

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 249)

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.

TITLE

Generation and analysis of 280,000 human expressed sequence tags

GENOME
MEDLINE

Genome Res. 6 (9), 807-828 (1996)

COMMENT

Contact: Wilton RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 339 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham.

FEATURES

source

1. .249

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="GDB:3919964"

/db_xref="taxon:9606"

/clone="IMAGE:530453"

/clone_lib="Stratagene fibroblast (#937212)"

/lab_host="SOLR cells (kanamycin resistant)"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. W138 cell line. Average insert size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT

73 a 59 c 43 g 74 t

ORIGIN

Query Match

0.5%; Score 22; DB 21; Length 249;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 tatccaccttttgggaagcag 950

|||||

Db 55 TATCCACCTTTTGGGAAGCAG 76

RESULT 13


```

AA365447
LOCUS EST76221 Pineal gland II Homo sapiens cDNA 5' end, mRNA sequence. 21-APR-1997
ACCESSION AA365447
VERSION AA365447.1 GI:2017767
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 253)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Whai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Frichman,J.L., Geoghagen,N.S.,
Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dumke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M., and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
12140200
Other_ESTs: THC179592
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018690556
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi.html)
Seq primer: M3 Reverse.
FEATURES
source Location/Qualifiers
1..253
/organism="Homo sapiens"
/db_xref="ATCC (inhost):170174"
/db_xref="taxon:9606"
/clone_lib="Pineal gland II"
/dev_stage="adult"
/note="Organ: pineal body; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 67 a 64 c 68 g 49 t 5 others
ORIGIN
Query Match 0.5%; Score 22; DB 25; Length 253;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 929 tatccacctttggggaagcag 950
|||||
Db 220 TATCCACCTTTGGGGAAGCAG 241
RESULT 14
AA876973/c
LOCUS EST76973 266 bp mRNA EST 25-MAR-1998
DEFINITION n49c05.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1275080,
mRNA sequence.
ACCESSION AA876973
VERSION AA876973.1 GI:2986050
QY 929 tatccacctttggggaagcag 950
|||||
Db 220 TATCCACCTTTGGGGAAGCAG 241
Query Match 0.5%; Score 22; DB 25; Length 253;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 929 tatccacctttggggaagcag 950
|||||
Db 220 TATCCACCTTTGGGGAAGCAG 241
Query Match 0.5%; Score 22; DB 32; Length 266;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 929 tatccacctttggggaagcag 950
|||||
Db 240 TATCCACCTTTGGGGAAGCAG 219
RESULT 15
AA860355/c
LOCUS EST7604.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394575
DEFINITION 3', mRNA sequence.
ACCESSION AA860355
VERSION AA860355.1 GI:2954350
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

```

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 266)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: 40ml3 fwd. Et from Amersham.
FEATURES
source Location/Qualifiers
1..266
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1275080"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from metastatic prostate
lesion of the bone, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."
BASE COUNT 73 a 46 c 59 g 88 t
ORIGIN
Query Match 0.5%; Score 22; DB 32; Length 266;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 929 tatccacctttggggaagcag 950
|||||
Db 240 TATCCACCTTTGGGGAAGCAG 219
RESULT 15
AA860355/c
LOCUS EST7604.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394575
DEFINITION 3', mRNA sequence.
ACCESSION AA860355
VERSION AA860355.1 GI:2954350
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

```

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 742 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 256.

FEATURES

source

1. .271
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1394575"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7m3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTCACATCTGAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7m3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
78 a 43 c 56 g 94 t
ORIGIN

Query Match 0.5%; Score 22; DB 32; Length 271;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 929 tatccaccttttgggaagcag 950
|||||
Db 234 TATCCACCTTTGGGGAAGCAG 213

Search completed: September 3, 2000, 17:40:16
Job time: 17852 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 17:39:54 ; Search time 2736.15 Seconds
(without alignments)
5255.724 Million cell updates/sec

Title: US-09-183-972-3
Perfect score: 3261
Sequence: 1 taacacaagaaggttactct.....tactatgacataatcaat 3261

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 5247842 seqs, 2204914090 residues
Word size : 0
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*

- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_est45:*
- 80: gb_est46:*
- 81: gb_est47:*
- 82: em_est35:*
- 83: em_est36:*
- 84: em_est37:*
- 85: gb_est48:*
- 86: gb_est49:*
- 87: gb_est50:*
- 88: gb_est51:*
- 89: gb_est52:*
- 90: gb_est53:*
- 91: gb_est54:*
- 92: gb_est55:*
- 93: gb_gss1:*
- 94: gb_gss2:*
- 95: gb_gss3:*
- 96: gb_gss4:*
- 97: em_gss1:*
- 98: em_gss2:*
- 99: em_gss3:*
- 100: em_gss4:*
- 101: gb_gss5:*
- 102: gb_gss6:*
- 103: gb_gss7:*
- 104: gb_gss8:*
- 105: gb_gss9:*
- 106: em_gss5:*
- 107: em_gss6:*
- 108: em_gss7:*
- 109: em_gss8:*
- 110: em_gss9:*
- 111: em_gss10:*
- 112: em_gss11:*
- 113: gb_gss10:*
- 114: gb_gss11:*
- 115: em_gss12:*
- 116: gb_gss12:*

QY 1442 atggcctacacctccctgtcagaagctccacctttttttatggatcaagcatctctct 1501
 |||||
 Db 181 ATGGCTCTACCTCCCTGTGTCAGAGCTCCACCTTCTTTATGGCATCAAGCATCTCTCT 240
 |||||
 QY 1502 ctgactgatcaaggaccacagacatacaatggccactgaccagacaaatgctagtaccagg 1561
 |||||
 Db 241 CTGACTGATCAAGGCACACAGATACAATGCCACTGCCAGACAATGCTAGTACCAGG 300
 |||||
 QY 1562 ctcacatcccc 1573
 |||||
 Db 301 CTCACCATCCCC 312
 |||||

RESULT 2
 W26960/c 561 bp mRNA EST 08-MAY-1996
 LOCUS 16h10 Human retina cDNA randomly primed sublibrary Homo sapiens
 DEFINITION cDNA, mRNA sequence.
 ACCESSION W26960
 VERSION W26960.1 GI:1306188
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 561)

AUTHORS Macke, J., Smallwood, P. and Nathans, J.
 TITLE Adult Human Retina cDNA
 JOURNAL Unpublished (1996)
 COMMENT On May 10, 1995 this sequence version replaced gi:805808.

Contact: Dr. Jeremy Nathans
 Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics
 Johns Hopkins School of Medicine
 725 North Wolfe Street, Baltimore, MD 21205
 Tel: 410 955 4678
 Fax: 410 614 0827
 Email: jeremy.nathans@gmail.bs.jhu.edu
 Clones from this library are NOT available.

PCR Primers

FORWARD: CTTTGTGACCAAGTTCAGCCTGGTAAAGT

BACKWARD: GAGGTGGCTTATGATATTTCTTCAGGGTAA

Seq primer: GGGTAAAGCAAAAGAAAT.

Location/Qualifiers

1. .561

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Human retina cDNA randomly primed sublibrary"

/sex="mixed (males and females)"

/tissue_type="retina"

/dev_stage="adult"

/lab_host="E. coli strain K802"

/note="Organ: eye; Vector: lambda gt10; Site:1: EcoRI;
 Site:2: EcoRI; The library used for sequencing was a
 sublibrary derived from a human retina cDNA library.
 Inserts from retina cDNA library DNA were isolated,
 randomly primed, PCR amplified, size-selected, and cloned
 into lambda gt10. Individual plaques were arrayed and
 used as templates for PCR amplification, and these PCR
 products were used for sequencing."

129 a 110 c 152 g 138 t 32 others

BASE COUNT

ORIGIN

Query Match 8.2%; Score 266; DB 91; Length 561;

Best Local Similarity 99.3%; Pred. No. 2.4e-132;

Matches 416; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1226 caatcttgatggtgggacaattcagttcactgatgaattgtgtgactccagcc 1285
 |||||

Db 465 CAATCTTGATGTTGGGACAAATTCAGTTTCACTGATGAATTCGTGTCATCCAGCC 406
 |||||

QY 1286 ttgtgctgacacccaatcagatgctccacatcttttctgttataacagagatgct 1345

Db 405 TTTGCTCTGACACCCCAATCAGANCTGCCACATCTTTTGTGTATACAGAGATGCT 346
 |||||
 QY 1346 actttgactcagaactcctcctgttgaaacccagcttgagacagtgagacagcagag 1405
 |||||
 Db 345 ACTTTGAGTCAGAAATCTTCTCTGTTGAACCCAGCTTGAGACAGTGGACGGACAGAG 286
 |||||
 QY 1406 catggtctacctgacactcttggctccacctgtatggcctctacctccctgtcagaa 1465
 |||||
 Db 285 CATGGCTACTGACACTTCTTGGTCTCCACCTGCTATGGCTCTACCTCCTCTGTGAGNA 226
 |||||
 QY 1466 gctccacctttttatggcatcaagcatcttctctgtactgatcaaggcaccagat 1525
 |||||
 Db 225 GCTCCACCTTTCTTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGCCACACAGAN 166
 |||||
 QY 1526 acaatggcactgaccagacaaatgctagtaccagggtccacatccaccacagtattat 1585
 |||||
 Db 165 ACAATGGCCTGACGACAGCAATGCTAGTACAGGGCTCAGCATCCACCATGATTAT 106
 |||||
 QY 1586 tctgcaatcagccaaactggtctggaattccacatccacctgcactctcagatgacag 1644
 |||||
 Db 105 TCTNCAATCAGCCAACTGCTGGGAATTTCAATCCACCTGCAATCTTCAGATGACAG 47
 |||||

RESULT 3

AA296278

LOCUS

EST10795 Umbilical vein endothelial cells II Homo sapiens cDNA 5'

DEFINITION end, mRNA sequence.

ACCESSION AA296278

VERSION AA296278.1 GI:1948653

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 298)

AUTHORS

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,

Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,

White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wei, C.,

Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,

Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,

Glocke, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S.Jr.,

Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,

Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, X.,

Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,

He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,

Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H.,

Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M.,

Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseitine, W.A., Fields, C.,

Fraser, C.M., and Venter, J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL

MEDLINE

COMMENT

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavetigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1. .298

/organism="Homo sapiens"

/db_xref="ATCC (inhost):194417"

FEATURES

source


```

/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCI_CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (cloneIDs 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaudo."
BASE COUNT      105 a      89 c      62 g      131 t
ORIGIN

Query Match      0.7%; Score 22; DB 42; Length 387;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 tcaaggcaccacagatacaatg 1531
|||||
Db 310 TCAAGGCACCACAGATACATG 289

RESULT 8
AI695882/c      391 bp      mRNA      EST      14-DEC-1999
LOCUS      ts87b06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2238227 3',
DEFINITION      mRNA sequence.
ACCESSION      AI695882
VERSION      AI695882.1 GI:4983782
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 391)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Jun 5, 1998 this sequence version replaced gi:3189232.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaudo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrr/image/image.html
Insert Length: 477 Std Error: 0.00
Seq primer: -400P from Gibco.
Location/Qualifiers
1. .391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2238227"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCI_CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (cloneIDs 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaudo."
FEATURES
source

```

```

BASE COUNT      106 a      90 c      63 g      132 t
ORIGIN

Query Match      0.7%; Score 22; DB 43; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 tcaaggcaccacagatacaatg 1531
|||||
Db 310 TCAAGGCACCACAGATACATG 289

RESULT 9
AI093483/c      446 bp      mRNA      EST      18-AUG-1998
LOCUS      qd15b03.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION      IMAGE:1696301 3' similar to contains MER18.b3 MER18 repetitive
element ;, mRNA sequence.
ACCESSION      AI093483
VERSION      AI093483.1 GI:3432459
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 446)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Jun 5, 1998 this sequence version replaced gi:3188975.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 445.
Location/Qualifiers
1. .446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1696301"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: uterus; Vector: pT7T3-Pac; Site:1; Not I;
Site:2; Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AAGTGGAGAAATTCGGCGCCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaudo."
BASE COUNT      111 a      101 c      80 g      154 t
ORIGIN

Query Match      0.7%; Score 22; DB 35; Length 446;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 tcaaggcaccacagatacaatg 1531
|||||
Db 312 TCAAGGCACCACAGATACATG 291

RESULT 10
AW303461/c      503 bp      mRNA      EST      18-JAN-2000
LOCUS      AW303461

```

DEFINITION xv19d01.xl Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2813569 3' similar to contains Alu repetitive
element; contains MER18.B3 MER18 repetitive element ;, mRNA
sequence.
AW303461
VERSION AW303461.1 GI:6713150
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 503)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 15, 1998 this sequence version replaced gi:3225011.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: 40UP from Gibco
High quality sequence stop: 463.
Location/Qualifiers
FEATURES
source
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2813569"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 114 a 116 c 96 g 177 t
ORIGIN

Query Match 0.7%; Score 22; DB 70; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1510 tcaaggcaccacagatacaatg 1531
|||||
Db 313 TCAAGGCACCACAGATACAATG 292

RESULT 11
LOCUS AQ871114 456 bp DNA GSS 03-NOV-1999
DEFINITION nbeb0042A08f CUGI Rice BAC Library (EcoRI) Oryza sativa
genomic clone nbeb0042A08f, genomic survey sequence.
ACCESSION AQ871114

VERSION AQ871114.1 GI:6221565
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 456)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)

COMMENT On Feb 19, 1999 this sequence version replaced gi:4146913.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: wing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 36
High quality sequence stop: 411.
Location/Qualifiers
FEATURES
source
1..456
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbeb0042A08f"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocytledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT 137 a 70 c 70 g 178 t 1 others
ORIGIN

Query Match 0.6%; Score 21; DB 116; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3168 aaaaattctactttctgtt 3188
|||||
Db 46 AAAAATCTACTTTCTGT 66

RESULT 12
LOCUS AW263637 606 bp mRNA EST 28-DEC-1999
DEFINITION xh81b12.xl Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2700863 3' similar to SM:NUPL_HUMAN P52594 NUCLEOPORIN-LIKE
PROTEIN RIP ;, mRNA sequence.
ACCESSION AW263637
VERSION AW263637.1 GI:6640453
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 606)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 418.

FEATURES
source

1..606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2700863"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI.CGAP.GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
228 a 109 c 89 g 180 t

Query Match 0.6%; Score 21; DB 70; Length 606;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2409 ttaaaagtccaaaatcaaca 2429
|||||
DB 520 TTAAGAAGTCCAAAATCAACA 540

RESULT 13

AO284104
LOCUS 164 bp DNA GSS 27-APR-1999
DEFINITION RPC11-80H22-TV RPC1-11 Homo sapiens genomic clone RPC1-11-
80H22, genomic survey sequence.
ACCESSION AO284104

VERSION AO284104.1 GI:3910422
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164)

REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
AUTHORS Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.

TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)

COMMENT Other_GSSs: RPC11-80H22.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7

Class: BAC ends.

FEATURES Location/Qualifiers
source 1..164
/organism="Homo sapiens"
/db_xref="GDB:7530525"
/db_xref="taxon:9606"
/clone="RPC1-11-80H22"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
30 a 38 c 30 g 66 t

Query Match 0.6%; Score 20; DB 96; Length 164;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 ttttggtttttggatttt 172
|||||
DB 46 TTTTGGTTTGGATTTT 65

RESULT 14
LOCUS AO264725 228 bp DNA GSS 27-OCT-1998
DEFINITION CITBI-El-2506F18-TF CITBI-El Homo sapiens genomic clone
2506F18, genomic survey sequence.
ACCESSION AO264725

VERSION AO264725.1 GI:3792925
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 228)

REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
AUTHORS Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.

TITLE Use of a random human BAC End Sequence database for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished (1998)

Other_GSSs: CITBI-El-2506F18.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21

Class: BAC ends.

FEATURES Location/Qualifiers
source 1..228
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2506F18"
/clone_lib="CITBI-El"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
38 a 57 c 51 g 82 t

BASE COUNT
ORIGIN

Query Match 0.6%; Score 20; DB 96; Length 228;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 16:21:22 ; Search time 6962.07 Seconds
(without alignments)
835.943 Million cell updates/sec

Title: US-09-183-972-3
Perfect score: 3261
Sequence: 1 taacacgaaggttactct.....tactatatgacataatcaat 3261

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 972840 seqs, 892348106 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: em_fun.*
17: em_hum1.*
18: em_hum2.*
19: em_in.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_sy.*
29: em_un.*
30: em_vi.*
31: gb_htg1.*
32: gb_htg2.*
33: gb_in1.*
34: gb_in2.*
35: em_ba1.*
36: em_ba2.*
37: em_hum3.*
38: em_hum4.*
39: gb_pr4.*
40: gb_htg3.*
41: gb_htg4.*
42: gb_htg5.*
43: gb_htg6.*

- 44: gb_htg7.*
45: em_htg1.*
46: em_htg2.*
47: em_htg3.*
48: em_hum5.*
49: gb_pl3.*
50: gb_pr5.*
51: gb_htg8.*
52: gb_htg9.*
53: gb_htg10.*
54: gb_htg11.*
55: gb_htg12.*
56: gb_htg13.*
57: gb_htg14.*
58: gb_in3.*
59: gb_htg15.*
60: gb_htg16.*
61: gb_htg17.*
62: em_htg4.*
63: em_htg5.*
64: em_htg6.*
65: em_htg7.*
66: em_hum6.*
67: gb_htg18.*
68: gb_htg19.*
69: gb_htg20.*
70: gb_htg21.*
71: gb_htg22.*
72: gb_htg23.*
73: gb_htg24.*
74: gb_htg25.*
75: gb_htg26.*
76: gb_htg27.*
77: gb_htg28.*
78: gb_htg29.*
79: gb_htg30.*
80: gb_htg31.*
81: gb_v11.*
82: gb_v12.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2619	80.3	3268	39	AF047492	AF047492 Homo sapi
2	653	20.0	1235	11	HSIMPG17	AF017776 Homo sapi
3	534	16.4	816	11	HSIMPG13	AF017772 Homo sapi
4	534	16.4	194704	32	AL157379	AL157379 Homo sapi
5	249	7.6	444	11	HSIMPG10	AF017769 Homo sapi
6	236	7.2	537	11	HSIMPG02	AF017761 Homo sapi
7	236	7.2	194704	32	AL157379	AL157379 Homo sapi
8	170	5.2	448	11	HSIMPG14	AF017773 Homo sapi
9	169	5.2	422	11	HSIMPG03	AF017762 Homo sapi
10	153	4.7	477	11	HSIMPG15	AF017774 Homo sapi
11	141	4.3	377	11	HSIMPG07	AF017766 Homo sapi
12	112	3.4	555	11	AF047491	AF047491 Macaca fa
13	109	3.3	311	11	HSIMPG06	AF017765 Homo sapi
14	81	2.5	438	11	HSIMPG12	AF017771 Homo sapi
15	79	2.4	347	11	HSIMPG11	AF017770 Homo sapi
16	78	2.4	310	11	HSIMPG01	AF017760 Homo sapi
17	67	2.1	304	11	HSIMPG05	AF017764 Homo sapi
18	62	1.9	300	11	HSIMPG08	AF017767 Homo sapi
19	43	1.3	380	11	HSIMPG16	AF017775 Homo sapi
20	33	1.0	300	11	HSIMPG04	AF017763 Homo sapi
21	24	0.7	84769	55	AC008800	AC008800 Homo sapi
22	24	0.7	123589	55	AC008796	AC008796 Homo sapi
23	24	0.7	157249	69	AC025180	AC025180 Homo sapi
24	24	0.7	162013	71	AC019347	AC019347 Homo sapi


```
25 24 0.7 181743 55 AC008823 Homo sapi
26 23 0.7 294 11 HSIMPC09
27 23 0.7 187117 70 AC009686 Homo sapi
28 22 0.7 4695 11 HS0802204
29 22 0.7 109667 42 AC015789 Homo sapi
30 22 0.7 133176 73 AC009058 Homo sapi
31 22 0.7 140266 32 AC0138726 Homo sapi
32 22 0.7 155652 43 AC009687 Homo sapi
33 22 0.7 160724 73 AC009024 Homo sapi
34 22 0.7 164148 43 AF215844 Homo sapi
35 22 0.7 170407 56 AC011177 Homo sapi
36 22 0.7 173858 32 AL137246 Homo sapi
37 22 0.7 196010 72 AC019124 Homo sapi
38 22 0.7 256271 76 AC048337 Homo sapi
39 22 0.6 120666 67 AC010454 Homo sapi
40 21 0.6 155941 78 AC023757 Homo sapi
41 21 0.6 161849 43 AC012254 Homo sapi
42 21 0.6 167237 39 AC007938 Homo sapi
43 21 0.6 175433 31 AP001547 Homo sapi
44 21 0.6 183414 43 AC011214 Homo sapi
45 21 0.6 185043 72 AC018964 Homo sapi
```

ALIGNMENTS

```
RESULT 1
LOCUS AF047492 3268 bp mRNA PRI 26-OCT-1999
DEFINITION Homo sapiens interphotoreceptor matrix proteoglycan 150 (IMPGL)
ACCESSION AF047492
VERSION AF047492.2 GI:6118565
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Kuehn,M.H. and Hageman,G.S.
TITLE Expression and characterization of the IPM 150 gene (IMPGL)
product, a novel human photoreceptor cell-associated
chondroitin-sulfate proteoglycan
JOURNAL Matrix Biol. 18 (5), 509-518 (1999)
MEDLINE 20068045
REFERENCE 2 (bases 1 to 3268)
AUTHORS Kuehn,M.H. and Hageman,G.S.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1998) Ophthalmology and Visual Sciences,
University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA
52240, USA
REFERENCE 3 (bases 1 to 3268)
AUTHORS Kuehn,M.H. and Hageman,G.S.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1999) Ophthalmology and Visual Sciences,
University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA
52240, USA
REMARK Sequence update by submitter
COMMENT On Oct 26, 1999 this sequence version replaced gi:2906231.
FEATURES
source
1. .3268
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="between markers D6S280 and D6S1659"
/tissue_type="retina"
1. .3268
/gene="IMPGL"
131..2524
/gene="IMPGL"
/note="IPM 150"
/codon_start=1
/product="interphotoreceptor matrix proteoglycan 150"
```

```
/protein_id="AAC03789.2"
/db_xref="GI:6118566"
/translaton="MYLEPRAIFVFWIFLQVQGTDKDISINIIYHSETKIDNPPRNET
TESTEKMYKMTMRIRFDLAKHRTKSAFFPTGVKPOESMKOILDSLOAYVRLVC
QEAWEAYRIFLDRIPTDGETYODWISICOQETKFLDIGNESNSOEHLDLQORIKO
RSPDRKDISAEKILGEGEIVISTDVANSLGFPPTPDDTLNEILDNLNTK
MPTEKETFAVLEQORVSLNFKRAELADQSQYIOELAKSLOQMKIFKK
LPGFKRIHYLGFPRKRDGSSSTNQLTAIFKRHSAEASPAISLSDSKNIESE
VYHGTMEEDKQEIYLTATDLKRLISKALEEESQSLDVGITQFTDETAGSLPAGPTQ
SELPTFAVITEDATLSPELPPVEPOLETVDGAHGLPDTSPSPAMASTSLSEAPPF
FWASSIFSLTDQGTDTMATDQMLVPGLTIPISDYSAISQLALGISHPPASSDDRS
SAGGEDVHRLDEMDLSDPAPSELSSEIVSPDHFLEDTTPVSAHQITITSSMTI
APKRGELVFFSLRVANMAFSNDLFNKSLEYRALEQFTQLLYLRSLNLTGFKOLE
ILNFRGSIYVNSKMKFAKSVNTLTAVHGLVEDFRSAQAQOLHLEISYLSINIPA
DQADPKFLACGEFAOCVKNRTEAEACKPGYDSQSLDGLPEGLCGPGTECEVL
OGKGAFCRLPDHSENOAYKTSVKKFNQNNKVISKRNSSELLTVEVEEFHQDMEGN"
BASE COUNT 1044 a 724 c 664 g 836 t
ORIGIN
Query Match 80.3%; Score 2619; DB 39; Length 3268;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3149; Conservative 0; Mismatches 1; Indels 4; Gaps 4;
QY 112 atgcacaaagccaggaatgtatttggaaactagaagagctattttttttttggattt 171
Db 115 ATGCACAAAAGCCAGATGTATTGGAACTAGAAAGAGCTATTATTTTGTGATTTT 174
QY 172 tctccaaagttcaaggaaactaaagatatctccattaaacatatataccattctgaaataaga 231
Db 175 TCTCCAAAGTTCAAGGAACCAAGATATCTCCATTAAATATACCATCTGAAACTAAGA 234
QY 232 catagacaatcccccaagaaatgaacaactgaagactgaagaaatgtacaaaatgtc 291
Db 235 CATAGACAATCCCCCAAGAAATGAACAACACTGAAAGTACTGAAAAAATGTACAAAATGTC 294
QY 292 aactatgagcgaatatttcgatttggcaagcatcgacaaaagatcgcaattttttccc 351
Db 295 AACTATGAGAGCAATATTCGATTGGCAAGCATCGAACAAAAGATCCGCATTTTCCC 354
QY 352 aacgggggttaagctgtgtccacaggaatccatgaacagattttagacagcttccaagc 411
Db 355 AAGCGGGGTTAAAGTGTGTCCACAGGAATCCATGAAACAGATTTTAGACAGCTCTCAAGC 414
QY 412 ttattatagattgagagtggtcaggaagcagtgatggaagcatatcgatctttcttggg 471
Db 415 TTATTATAGATTGAGAGTGTTGTCAGGAAGCAGTATGGAAGCATATCGGATCTTCTGGA 474
QY 472 tcgcatccttgacacaggggaatatacgagctgggtgggtcagcatctgccagcaggagacctt 531
Db 475 TCGCATCCTTGACACAGGGGAATATCAGGACTGGGTGAGCATCTGCCAGCAGGAGACCTT 534
QY 532 ctgctctttgacatttggaataacttcagcaattccacaggagcacttgatcttctcca 591
Db 535 CTGCTCTTTGACATTTGGAAAAAATCTTCAGCAATTCCTCAGGAGCACCTGGATCTCTCCA 594
QY 592 gcagagaataaacacagagaagtttccctgcagagaaagatgaaatatctgcagagaagac 651
Db 595 GCAGAGAAATAAACAGAGAAGTTTCCCTGCACAGAAAAGATGAAATATCTGCAGAGAAGAC 654
QY 652 attggggagagcctgtgtgaaccattgtcatttcaacagatgttgcacagctctcactgg 711
Db 655 ATGGGGAGAGCCTGGTGAACCATTTGTCATTTCACAGATGTTGCCAACGCTCTCATTGG 714
QY 712 gcctttcccttactcctctgatgacacccctctcctcaatgaaattctcgataatacactcaa 771
Db 715 GCCTTTCCCTCTACTCTCTGATGACACCCCTCTCTCAATGAAATTCCTGATAATACACTCAA 774
QY 772 cgacaccaagatgctctacacagaaaagagaagaattcgctgtgttgaggaggagag 831
Db 775 CGACACCAAGATGCTCTACAACAGAAAGAGAAACAGAAATTCGCTGTGTGGAGGAGAGAG 834
QY 832 ggtggagctcagcgtctctctctgtgtaaacccagaagtccaaggcagagctcgctgactccca 891
```

Db 885 |||||GCTGAGCTACGCGTCTCTCTGTTAAACAGAAAGTTCAAGGCGAGAGCTCGCTGACTCCCA 894
Qy 892 gTccccatattaccaggagctagcaggaagtcaccaacttcagatgcgaagatatttaa 951
Db 895 GTCCCCATATTACCAGGAGCTAGCAGGAAGTCCCAACTTCAGATGCAGAAAGATATTAA 954
Qy 952 gaaacttccaggattcaaaaaatccatgtgttagatttagaccnaaagaaaaaga 1011
Db 955 GAAACTTCAGGATTCAAAAAATCCATGTGTAGGATTTAGACCAAGAAAGAAAGA 1014
Qy 1012 tggctaaagctccagagagatgcaacttaaggccatcttttaagagacacagtgcaagc 1071
Db 1015 TGGCTCAAGCTCCACAGAGATGCAACTTACGGCCATCTTTAAGAGACACAGTGCGAAGC 1074
Qy 1072 aaaaagccctgaagtgacctctgtctttttagttccaaacaaatgaaagtgaagaagt 1131
Db 1075 AAAAGCCCTGCAAGTGACCTCTGTCTTTTGATTCCAAACAAATTCGAAATGAGGAAAGT 1134
Qy 1132 ctatcatgaaacaaatggaggaggaagcaagcaaccagaaatctatctcacagctcacagcct 1191
Db 1135 CTATCATGGAACCATGGAGGAGACAAACACAGAAATCTATCTCACAGCTACAGACCT 1194
Qy 1192 caaaaggctgatacagaagaagcactagaggaagaaacaatctttggatgtggggacaattca 1251
Db 1195 CAAAAGGCTGATCAGCAAAAGCACTAGAGGAAGAACAAATCTTTGGATGTGGGGCAATTC 1254
Qy 1252 gttcaactgaataattgtggatcaactgccagcctttgtctcgacacccaatcagagct 1311
Db 1255 GTTCACTGATGAAATGCTGGATCACTGCGAGCTTTGGTCTTGACACCCCAATCAGAGCT 1314
Qy 1312 gccacatctttgtgttaaacagagatgctactttgagtcagaaacttctctctgt 1371
Db 1315 GCCACATCTTTGCTGTTTAAACAGAGGATGCTACTTTGAGTCCAGAACTTCTCTCTGT 1374
Qy 1372 tgaacccagcttgagacagtgagcagagagcagatggtctaccttgacactcttggtc 1431
Db 1375 TGAACCCAGCTTGAGACAGTGAGCGAGGAGAGCATGGTCTACCTGACACTTCTTGCTC 1434
Qy 1432 tccactgtatggcctctacctccctgtcagaagctccacctttctttatggcataag 1491
Db 1435 TCCACCTGTATGGGCTCTACCTCCCTGTGCAAGAGTCCACCTTTCTTTTATGTCATCAG 1494
Qy 1492 catcttctctgactgaacaaggccacagatacaatggccactgacagacaatgct 1551
Db 1495 CATCTTCTCTGACTGATCAAGGCCACACAGATACAATGGCCACTGACCAGACAATGCT 1554
Qy 1552 agtacagggtcaccatccaccagctgattattctgcaatcagccaactggctctggg 1611
Db 1555 AGTACAGGGCTCACCATCCCCACACAGTGATTTCTGCAATCAGCCAACTGGCTCTGGG 1614
Qy 1612 aattccaatccacttcgactcttcagatgacagccgatacaagtgcaggtggcgaagatat 1671
Db 1615 AATTTACATCCACCTGCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGCGAAGATAT 1674
Qy 1672 ggtcagaacctagatgaatggatgtgtgcacatcctctgccccatctgaggtaccaga 1731
Db 1675 GGTACAGACCTTAGATGAAATGGATGTGTGACATCTCTGCCCCATCTCAGGTACCAGA 1734
Qy 1732 gctcagcgaatatgttctgtccagatcatcttctggaggataccactctgtctcagc 1791
Db 1735 GCTACGGAATATGTTCTCTGCCAGATCATTTCTTGGAGGATACCACTCCTGTCCTCAGC 1794
Qy 1792 tttaacagtataccactagttctctatgaccattggcccccaaggccgagagctggtagt 1851
Db 1795 TTTACAGTATATCACCACTAGTTCTATGACCATGCCCCCAAGGGCCGAGAGCTGTAGT 1854
Qy 1852 gttcttcagctgcgtgtgttctaaacatggcctcttccaaacgaacctgttcaacagagctc 1911
Db 1855 GTTCTTCACTCTCGGTGTTGCTAAACATGGCTTCTCCACAGCCTGTTCAACAAGAGCTC 1914
Qy 1912 tctgagatccagagctctctggagcaacaaattcaacagctctggttccatatctacagatc 1971

Db 1915 TCTGGAATACCGAGCTCTGGAGCAACAATTCACACAGCTGCTGGTTCCATATCTACGATC 1974
Qy 1972 caatcttacaggatttaagcaacttgaaataacttaacttcagaaacggagtgattgtt 2031
Db 1975 CAATCTTACAGGATTTAAGCAACTTGAAATACTTAATCTTCAGAAACGGAGTGTGATTGT 2034
Qy 2032 gaatagcaaaatgaagtgttctgaagctgtgcggtataaacctcccaagggtgtgcacgg 2091
Db 2035 GAATAGCAAAATGAAGTTTGTAAAGTCTGTCCGTATAAATCTCCACCAAGGCTGTGCACGG 2094
Qy 2092 ggtcttgaggatttctgtctgcagcccaacaactccatctggaataagacagta 2151
Db 2095 GGTCTTGAGGATTTCTGTTCTGCTGCAGCCCAACACTCCATCTGGAATATAGACAGTA 2154
Qy 2152 ctctctcaacattgaaccagctgatcaagcagatccctgcgaagttccctggcctgcggcga 2211
Db 2155 CTCTCTCAACATTGAACAGCTGATCAAGCAGATCCCTGCAAGTTCCTTGGCTGCGGGA 2214
Qy 2212 atttgcccaatgtgtaagaacgaacggaactggaagcggagtgctgctgcaaacccagg 2271
Db 2215 ATTTGCCCAATGTCTAAAGAACGAACGGACTGAGGAAGCGAGTGTCTGCTGCAAAACCCAGG 2274
Qy 2272 atatgaagccagccaggcgtgagcgtctggaaccagcctctctggtg-ctggcacaacaa 2330
Db 2275 ATATGACAGCAGGGAGCCTGGACGCTCTGGAACCAAGGCTCTGTGGCCCTGGCAGAAA 2334
Qy 2331 ggaatgcgaggtcctccagggaaaggaggtccatgc-gggtt-ccagatcactctgaaaa 2388
Db 2335 GGAATGCGAGGCTCTCAGGGAAAGGGAGCTCCATGCAAGTTCGAGATCACTCTGAANA 2394
Qy 2389 tcaagcatacaaaactagttt-aaaaagttccaaaaatacaaaaaatacaagggttaacag 2447
Db 2395 TCAAGCATACAAAACACTAGTGTAAAGAGTTCCAAAAATCAACAAAAATAACAAGGTAATCAG 2454
Qy 2448 taaaagaattctgaattactgacgtagaatatgaagaatttaaccatacgaattggga 2507
Db 2455 TAAAGAAATTTCTGAATTTACTGACCGTAGAATATGAAGAAATTTAACCATCAAGATTTGGGA 2514
Qy 2508 aggaataaaaaactgaaatgtacaattatcacttaggtctatctcaagagagatgattt 2567
Db 2515 AGGAAATTAANAACGTGAAATGTACAAATATCACTTAGGCTATCTCAAGACAGATGATT 2574
Qy 2568 gccttctcaagggaataatggagacagggcatattcatgggtcatcaaaaaatccagacatacag 2627
Db 2575 GCCTTCTCAAGGAAATGGAGACAGGCATATTCATGGGTCAATAAATCCAGACATACAG 2634
Qy 2628 tcaacactgagaatcagcacacacacatttccaaatatagaagagtcagtacttgcaaa 2687
Db 2635 TCAACACTGAGAAATCAGCACACACACATATTTCAAAATATAGAAGAGTCACTGACTTGGCAA 2694
Qy 2688 ccagtaattctgaaaaaaaagacacttactattattataaaaccccaaatcaaatcagcg 2747
Db 2695 CCAGTAATTTCTGAAAAAAGACACTTACTTATTTAAAAACCCCAATGCAATCAGCG 2754
Qy 2748 aaacataatttactattcttgatgtagtcaaaatgatcataagccaggtttgctccc 2807
Db 2755 AAACATATTTTACTATTTCTGGATGATAGTCAAAAATGATCATAGCCAGGTTGCTTCC 2814
Qy 2808 accttccctgaaaaattttactcacagatcatttgcaacaagcatagcttactattgttt 2867
Db 2815 ACCTTCCCTGAAAAATTTTACTTCACAGATCAATTCGAACAACAGCATAGCTTACTTATTGTTT 2874
Qy 2868 agggactgaacaaatttatgggaagcaaacctcttatatgtagaagaacatattaaaag 2927
Db 2875 AGGACTGAACAAATTTATTTGGGAAGCAAACTCTTTATATGCTAGAAAAGTACATTTAAAAG 2934
Qy 2928 atgaactactacgagggagatgcaggtctctctaaacgaatgaatgtatgtatgtgta 2987
Db 2935 ATGACTACTTACGAGGGAGATGCGAGTCTCTTAAACGCGCATGAATGTATGTATGTGTGTA 2994
Qy 2988 ggcaactagtgagtgatataatgctccacactcgctctgataaacaacaacccctcagtat 3047
Db 2995 GGCACCTGATGAGTGTATATATGCTCCACACTACGTCTGTATTAACACAAACCTTCAGTAT 3054

QY 3048 tcagttattagcacactagttttatatacgcactactcttaccatagtagactgtttgt 3107
|||||
Db 3055 TCAGTTATTAGGCACACTAGTTTATATAGCAACTACTGCTTACATAGTAGAGTGTGT 3114
|||||
QY 3108 tgccaataatttgaattgtcttcttaaaagaactgaggttcagatacacataccatgg 3167
|||||
Db 3115 TGCCAATAAATCTTGAATGTCTTCTTAAAGAAAGTGAAGTTCAGATACACATACCATGG 3174
|||||
QY 3168 aaaaatttacttcttcttactacacaaagctattttaagaagatgctatgttgga 3227
|||||
Db 3175 AAAAATCTTACTTCTTGTACTACACAAAGCTATTTTAAAGAGATGCTATGTGGGA 3234
|||||
QY 3228 gaaggggaaagtgttactatatacacaataatcaat 3261
|||||
Db 3235 GAAGGGCGAAGTGTACTATATGACATATCAAT 3268
|||||

RESULT 2
HSIMPG17 1235 bp DNA PRI 28-OCT-1998
LOCUS HSIMPG17 Homo sapiens interphotoreceptor matrix gene (IPM150), exon 17 and
DEFINITION complete cds.
ACCESSION AF017776
VERSION AF017776.1 GI:3800731
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Feilbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
Schmid, M. and Weber, B.H.
TITLE Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate
for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81(1-2):17-24(1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 1235)
AUTHORS Gehrig, A., Feilbor, U., Kessel, R., Hunt, D.M., Maunene-Husels, I.E.
and Weber, B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choroidretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRI)
Unpublished
JOURNAL 3 (bases 1 to 1235)
AUTHORS Feilbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humanogenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES
source
1. 1235
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
join(AF017760.1:<1..223,AF017761.1:97..330,
AF017762.1:112..278,AF017763.1:82..110,
AF017764.1:107..171,AF017765.1:139..242,
AF017766.1:163..303,AF017767.1:96..154,
AF017768.1:153..173,AF017769.1:92..339,
AF017770.1:148..224,AF017771.1:271..349,
AF017772.1:144..676,AF017773.1:111..330,
AF017774.1:71..269,AF017775.1:94..166,413..490)
/gene="IPM150"
/product="interphotoreceptor matrix"
order(AF017760.1:<1..310,AF017761.1:1..537,
AF017762.1:1..422,AF017763.1:1..300,AF017764.1:1..304,
AF017765.1:1..311,AF017766.1:1..377,AF017767.1:1..300,
AF017768.1:1..294,AF017769.1:1..444,AF017770.1:1..347,
AF017771.1:1..438,AF017772.1:1..816,AF017773.1:1..448,

AF017774.1:1..477,AF017775.1:1..380,1..490)
/gene="IPM150"
join(AF017760.1:157..223,AF017761.1:97..330,
AF017762.1:112..278,AF017763.1:82..110,
AF017764.1:107..171,AF017765.1:139..242,
AF017766.1:163..303,AF017767.1:96..154,
AF017768.1:153..173,AF017769.1:92..339,
AF017770.1:148..224,AF017771.1:271..349,
AF017772.1:144..676,AF017773.1:111..330,
AF017774.1:71..269,AF017775.1:94..166,413..490)
/gene="IPM150"
/codon_start=1
/product="interphotoreceptor matrix"
/protein_id="AAC68835.1"
/db_xref="GI:3800731"
/translation="MYLETRRAIFVFWLQVGTGKDISINLYHSETKDIDNPPNET
TESTEKYKMTMRIFDLAKHRTARSFAFFGVKVCQESMKQLDLSQAYRURVC
QEAWEAYRIFLDRIPTDEYQDWVSIQOQFCLFDIGKNFSQEHLDLLQRIKQ
RSFDRKDEISAERTLGPGETIVISDVANVSLGFPPLTDDTLNLDLNTDK
MPTRETEREFAVLEQREVELSVLNQKFAELADSQPYOELAGKSLQMQKIFKK
LPGFKIHVLGPRKPKDSSSEMOLTAIFKRHSARAKSPASLLSFDNSKIFSEE
VYHGTMEQKQEIYLTATDLKRLISKALEEQSLDVGTIQTDIASLSPAGPDTQ
SELPTSFAVITEDATLSPELPVPEQLETVDGAERGLDPTSWSPAMASTUSEAPPF
FMASIFSLTDQGTDDTMDQTLVPLGLTPTSDYSAISQALAGISHPPASDDSR
SAGDEDMVRHLDMDLSDTPAPSEVPELSEYVSPDHFLEDTTPYSALQYITTSMTI
APKRELVVFTSLRVANMAFNDFNKSLEYRALEQOFTQLLVYLSRNLTFKQLE
ILNFRNGSVIVNSKMKFASVPYNTKAVHGVLEDFRSAAQOLHLEIDSYSLNTEPA
DQADPCFLACGEFAQCVKNERTEAEACRCRPGYDSQSLDGLGLEGPGTKEEVL
QKGKAPCRLPDHSNQAYKTSVKFQNNQNNKVIKRNSELLTVEEFNHODWEGN"
413..490
/gene="IPM150"
BASE COUNT 414 a 216 c 234 g 371 t
ORIGIN
Query Match 20.0%; Score 653; DB 11; Length 1235;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 823; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 2438 aggttaacgttaaaagaattctgaattactacccatagaataatgaagaatttaaacatc 2497
Db 411 AGGTAATCAGTAAAAAGAAATTCGTATTACTGACCGTAGAATATGAAGAAATTTAACCATC 470
QY 2498 agattgggaagaaataaaactgaaatatacaattactcatttagctatctcaaga 2557
Db 471 AGATTGGGAAGAAATTAATAAACTGAAATGTACAAATATATCATTAGGCTATCTCAAGA 530
QY 2558 gagatgattgcttcttcaggaaaaatggagacagcattatcattgggtcatcaaatcc 2617
Db 531 GAGATGATTGGCTTCTCAAGGAAATGGACAGCGCATATTTCATGGTCATCAAAATCC 590
QY 2618 agacatacagtcacactgagaatcagcacacacacatttccaaatataagaagatcag 2677
Db 591 AGACATACAGTCAACACTGAGAATCAGACACACACCATATTTCAAATATATAGAAGAGTCATG 650
QY 2678 tacttggcaaccagtaaatcttgaa-aaaaagacacttacttattataaaacccaaa 2736
Db 651 TACTTGGCAACCAGTAAATCTTGAGAAAAAGACACTTACTTATTATTATAAAACCCCAA 710
QY 2737 tgcaatcagcgaacacatttttactattcttggatgatagtcacaaatgatcataagcca 2796
Db 711 TGCATACGCGAAACATATATTTTACTATTCTTGGGATAGTAGTCAAATATGATCATAGCCA 770
QY 2797 ggtttgcttccaccttccctgaaaaattttactcacagatcatttggcaacaagatagctt 2856
Db 771 GGTTCCTCCACCTTCCCTGAAATTTTACTCACAGATCATTTTGAACAAAGCATAGCTT 830
QY 2857 acttatgtttagggaactgaacaattatttgggaagacaaactctttatatgctagaagt 2916
Db 831 ACTTATTGTTAGGGACTGAACAAATTTATTGGGAAGCAAACTCTTTATATGCTAGAAAGT 890
QY 2917 acatttaagaatgactacttaccgagcgagatgcagggtctctctctaaacgcatgaatga 2976
|||||

```

Db 891 ACATTAAAGATGACTACTACGACGAGGAGATGCGAGGTCTCTCTAAACGCATGAATGTA 950
Qy 2977 tgaagtgttagcactgttagtgatgtatatgtccacactacgtctgataaacaca 3036
|||||
Db 951 TGTAGTGTGAGGACATGTAGTGTATATATGTCTCCACACTACGCTGTGATAACACA 1010
|||||
Qy 3037 aacctcagttattcagttatttaggcacactagttttttatcaccaactactgtttacatagta 3096
|||||
Db 1011 AACCTCAGTATTCAAGTATTAGGCACACTAGTTTATATACGCAACTACTGCTTACATAGTA 1070
|||||
Qy 3097 gactgtttgttgcataataattttgaattgttttttaaaagaactgagttcagatac 3156
|||||
Db 1071 GACTGTGTTGTGTCCTCAATATCTTTGAATTTGTTTAAAGAACTGAGGTTCCAGATAC 1130
|||||
Qy 3157 acatcacatggaatacttactttcttctgttactacacaaagctattttaaaagaactg 3216
|||||
Db 1131 ACATACCATGGAATAATCTTACTTTCTTGTCTACTACACAAAGCTATTTTAAAGAAGATG 1190
|||||
Qy 3217 ctatgttggaagagcggaagttgtactatatgacataatcaat 3261
|||||
Db 1191 CTATGTTGGAGAGGCGGAAGTGTACTATATGACATAATCAAT 1235
|||||

RESULT 3
HSIMPG13 --HSIMPG13 816 bp DNA PRI 28-OCT-1998
DEFINITION Homo sapiens Interphotoreceptor matrix gene (IPM150), exon 13
ACCESSION AF017772
VERSION AF017772.1 GI:3800727
KEYWORDS
SEGMENT
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 816)
Feibor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
Schmid,M. and Weber,B.H.
Genomic organization and chromosomal localization of the
Interphotoreceptor matrix proteoglycan-1 (IPM1) gene: a candidate
for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98356939
REFERENCE
1 (bases 1 to 816)
Gehrig,A., Feibor,U., Kessel,R., Hunt,D.M., Maumenee-Hussels,I.E.
and Weber,B.H.F.
Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRI)
Unpublished
JOURNAL
REFERENCE
3 (bases 1 to 816)
Feibor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.
Direct Submission
TITLE Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
JOURNAL
FEATURES
source
1. 816
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
144..676
/gene="IPM150"
BASE COUNT 200 a 207 c 154 g 255 t
ORIGIN
Query Match 16.4%; Score 534; DB 11; Length 816;
Best Local Similarity 100.0%; Pred. No. 2.9e-284;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1418 gacactctgtctccactgtgtgacctctacctccctcctgtcgaagaagctccaccttc 1477

```

```

Db 143 GACACTTTTGGTCTCCACCTGCTATGGCTCTACTCCCTGTGAGAAGCTCCACCTTTC 202
|||||
Qy 1478 ttatggcatcaagcatcttctctgactgatcaagcaccacacagatacaatggccact 1537
|||||
Db 203 TTTATGGCATCAAGCATCTTCTCTGACTGATCAAGCACCACAGATACATATGGCCACT 262
|||||
Qy 1538 gaccagacaatgttagtaccaggggtccaccatccccaccagtgattattctgcaatcagc 1597
|||||
Db 263 GACCAGACAATGCTAGTAGTACCAGGGCTCACCATCCCACCAGTGATTAATCTCAATCAGC 322
|||||
Qy 1598 caactggtctggaatttcacatccacctgcattcttcagatgacagcgatcaagtga 1657
|||||
Db 323 CAACTGGCTCTGGGAATTTTCATCCACTGCATCTTCAGATGACAGCGCATCAAGTGCA 382
|||||
Qy 1658 gttggcgaagatgttcagacacacatagataaattggatctgtcgcacctctgccccca 1717
|||||
Db 383 GGTGCGAAGATATGGTCAGACACCTAGATAAATGGATCTGCTCGACACTCCTGCCCA 442
|||||
Qy 1718 tctgaggtaccagagctcagcgaatatgttctgtctccagatcatttcttgagagatacc 1777
|||||
Db 443 TCTGAGGTACCAGAGCTCAGCGAATATGTTTCTGTCCCAGATCATTTCTTGAGGATACC 502
|||||
Qy 1778 actcgtctcagctttacagtatatcaccactagtcttatgacctgcccccaaggcc 1837
|||||
Db 503 ACTCTGTCTCAGCTTTACAGTATATACACCTAGTCTTATGACCATTTGCCCAAGGCG 562
|||||
Qy 1838 cgagagctggttagtcttctcagtcggtgttgtaacatggccttctccaaacacctg 1897
|||||
Db 563 CGAGAGCTGGTAGTGTCTTTCAGTCTGCTGTGTGTACATGGCCTTCTCCAAACACCTG 622
|||||
Qy 1898 ttcaacaagagctctctggagtaccgagctctggagcaacaattcacacagctg 1951
|||||
Db 623 TTCAACAGAGCTCTCTGAGTAGCGAGCTCTGGAGCAACAATTCACACAGCTG 676
|||||

RESULT 4
AL157379 194704 bp DNA HTG 22-MAR-2000
LOCUS Homo sapiens chromosome 6 clone RP1-62L18, *** SEQUENCING IN
DEFINITION PROGRESS ***, 28 unordered pieces.
ACCESSION AL157379
VERSION AL157379.2 GI:7159486
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 194704)
Sims,S.
Direct Submission
TITLE Submitted (22-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
JOURNAL
COMMENT On Mar 6, 2000 this sequence version replaced gi:7009540.
IMPORTANT: this sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00025 length: 6617bp
Contig_ID: 00126 length: 8477bp
Contig_ID: 00151 length: 1873bp
Contig_ID: 00275 length: 6849bp
Contig_ID: 00307 length: 1124bp
Contig_ID: 00337 length: 3565bp
Contig_ID: 00377 length: 5047bp
Contig_ID: 00383 length: 1165bp
Contig_ID: 00410 length: 2886bp
Contig_ID: 00414 length: 15535bp
Contig_ID: 00462 length: 1995bp
Contig_ID: 00501 length: 1028bp

```

Contig_ID: 00598 Length: 728bp
 Contig_ID: 00686 Length: 684bp
 Contig_ID: 00766 Length: 1115bp
 Contig_ID: 00821 Length: 12763bp
 Contig_ID: 00909 Length: 23204bp
 Contig_ID: 00911 Length: 7481bp
 Contig_ID: 00925 Length: 6805bp
 Contig_ID: 00932 Length: 1460bp
 Contig_ID: 00934 Length: 10010bp
 Contig_ID: 01026 Length: 2716bp
 Contig_ID: 01069 Length: 2283bp
 Contig_ID: 01121 Length: 10059bp
 Contig_ID: 01185 Length: 2031bp
 Contig_ID: 01186 Length: 1779bp
 Contig_ID: 01370 Length: 1640bp
 Contig_ID: 01386 Length: 1498bp.

* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 6617: contig of 6617 bp in length
 * 7417: contig of 800 bp
 * 15864: contig of 8447 bp in length
 * 15865: gap of 800 bp
 * 16665: 18537: contig of 1873 bp in length
 * 18538: 19337: gap of 800 bp
 * 19338: 26186: contig of 6849 bp in length
 * 26187: 26986: gap of 800 bp
 * 26987: 28110: contig of 1124 bp in length
 * 28111: 28910: gap of 800 bp
 * 28911: 32475: contig of 3565 bp in length
 * 32476: 33275: gap of 800 bp
 * 33276: 38322: contig of 5047 bp in length
 * 38323: 39122: gap of 800 bp
 * 39123: 40287: contig of 1165 bp in length
 * 40288: 41087: gap of 800 bp
 * 41088: 43973: contig of 2886 bp in length
 * 43974: 44773: gap of 800 bp
 * 44774: 60308: contig of 15535 bp in length
 * 60309: 61108: gap of 800 bp
 * 61109: 81103: contig of 19995 bp in length
 * 81104: 81903: gap of 800 bp
 * 81904: 82931: contig of 1028 bp in length
 * 82932: 83731: gap of 800 bp
 * 83732: 91019: contig of 7288 bp in length
 * 91020: 91819: gap of 800 bp
 * 91820: 98660: contig of 6841 bp in length
 * 98661: 99460: gap of 800 bp
 * 99461: 100575: contig of 1115 bp in length
 * 100576: 101375: gap of 800 bp
 * 101376: 114138: contig of 12763 bp in length
 * 114139: 114938: gap of 800 bp
 * 114939: 138142: contig of 23204 bp in length
 * 138143: 138942: gap of 800 bp
 * 138943: 146423: contig of 7481 bp in length
 * 146424: 147223: gap of 800 bp
 * 147224: 154028: contig of 6805 bp in length
 * 154029: 154828: gap of 800 bp
 * 154829: 156288: contig of 1460 bp in length
 * 156289: 157088: gap of 800 bp
 * 157089: 167098: contig of 10010 bp in length
 * 167099: 167898: gap of 800 bp
 * 167899: 170614: contig of 2716 bp in length
 * 170615: 171414: gap of 800 bp
 * 171415: 173697: contig of 2283 bp in length
 * 173698: 174497: gap of 800 bp
 * 174498: 184556: contig of 10059 bp in length
 * 184557: 185356: gap of 800 bp
 * 185357: 187387: contig of 2031 bp in length

* 187388 188187: gap of 800 bp
 * 188188 189966: contig of 1779 bp in length
 * 189967 190766: gap of 800 bp
 * 190767 192406: contig of 1640 bp in length
 * 192407 193206: gap of 800 bp
 * 193207 194704: contig of 1498 bp in length.

FEATURES
 Location/Qualifiers
 1..194704
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone_lib="62L18"
 /clone_lib="RPC1-1"
 BASE COUNT 53528 a 32672 c 33519 g 53383 t 21602 others
 ORIGIN

Query Match 16.4%; Score 534; DB 32; Length 194704;
 Best Local Similarity 100.0%; Pred. No. 2.3e-284;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1418 gacacttctgtctccactgctatggcctctacctcctgtcacaagctccaccttc 1477
 |||||
 Db 1761 GACACTTCTGTGGTCTCCACTGCTATGGCTCTACCTCCCTGTGAGAAGCTCCACCTTC 1820
 |||||
 QY 1478 ttatggcatcaagcatcttctctgactgacgacccacacacagatacaatggcact 1537
 |||||
 Db 1821 TTATGGCATCAAGCATCTCTCTGACTGATCAGGACACACATACAATGGCCACT 1880
 |||||
 QY 1538 gaccagacaatgctagtaccaggggtccacatccccaccagtgattatttgcacaagc 1597
 |||||
 Db 1881 GACCAGACAATGCTAGTACCAGGGCTCACCATCCCCACCAGTGATTATTCTGCAATCAGC 1940
 |||||
 QY 1598 caactgctctgggaatttccacatccacctgcatcttcagatgacagccgacataagtgca 1657
 |||||
 Db 1941 CAACTGGCTCTGGGAATTTACATCCACTGCATCTCTAGATGACAGCCGATCAAGTGCA 2000
 |||||
 QY 1658 ggTggcgaagatatggtcagacacccctagatgaaatgatatgtctgacacacctctgcccc 1717
 |||||
 Db 2001 GGTGGCGAAGATATGTTGTCAGACACCTAGATGAATGATGATGCTGTGACACTCTCTGCCCA 2060
 |||||
 QY 1718 tctgaggtaccagagctcagcgcaatgtttctgtcccagatcatcttcttggaggatacc 1777
 |||||
 Db 2061 TCTGAGGTACCCAGAGCTCAGCGAATATGTTCTGTCTCCAGATCATTTCTGGAGGATACC 2120
 |||||
 QY 1778 actctgtctcagctttacagtataccacactagttctctatgaccattgcccccaagggc 1837
 |||||
 Db 2121 ACTCTGTCTCAGCTTTACAGTATATCACCACTAGTCTTATGACCATTTGCCCAAGGGC 2180
 |||||
 QY 1838 cagagagctgtagtcttctcagctcgtgttgctaaacatggccttctccacgacctg 1897
 |||||
 Db 2181 CGAGAGCTGTGTAGTCTTCTTCAGTCTGGGTGCTAACATGGCTTCTCCACAGAGCTG 2240
 |||||
 QY 1898 ttcaacaagagctctctggagtagccagctctggagcaacaattcacacagctg 1951
 |||||
 Db 2241 TTCAACAAGAGCTCTCTGGAGTACCGAGCTCTGAGCAACAATTCACACAGCTG 2294
 |||||

RESULT 5

HSIMP10

LOCUS

Homo sapiens

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 444)

Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M., Schmid,M. and Weber,B.H.

AF017769.1

GI:3800724

10 of 17

human

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 444)

Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M., Schmid,M. and Weber,B.H.

28-OCT-1998

interphotoreceptor matrix gene (IPM150), exon 10.

444 bp

DNA

PRI

TITLE Genomic organization and chromosomal localization of the interphotoreceptor matrix proteoglycan-1 (IMPG1) gene: a candidate for 6q-linked retinopathies

JOURNAL Cytogetnet. Cell Genet. 81 (1), 12-17 (1998)

MEDLINE 98358139

REFERENCE 2 (bases 1 to 444)

AUTHORS Gehrig, A., Feilbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussels, I.E. and Weber, B.H.F.

TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150) localized to 6q14.2-q15 in autosomal dominant Stargardt-like macular dystrophy, progressive bifocal choriorretinal atrophy (PBCRA), and North Carolina macular dystrophy (MCDRI) Unpublished

JOURNAL 3 (bases 1 to 444)

REFERENCE Feilbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.

AUTHORS Direct Submission

TITLE Submitted (09-AUG-1997) Humanenetik, Universitaet Wuerzburg, Am Hubland, Wuerzburg D-97074, Germany

JOURNAL Location/Qualifiers

FEATURES source

1. .444

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="8"

/map="6q14.2-q15"

92. .339

/gene="IPM150"

BASE COUNT 133 a 86 c 101 g 124 t

ORIGIN

Query Match 7.6%; Score 249; DB 11; Length 444;

Best Local Similarity 100.0%; Pred. No. 3.6e-126;

Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 gctcaagctcagagatgcaacttaccgcccattttaaagagacacagtgagaagcaa 1073

|||||

Db 91 GCTCAAGCTCCACAGATGCAACTACGGCCATCTTTAAGAGACACAGTGCAGAAAGCAA 150

QY 1074 aaagccctgcaagtgcctcctcttcttgattcccaaaaaattgaaagtggagagctt 1133

|||||

Db 151 AAAGCCCTGCAAGTGACCTCCCTGCTTTGATTCACAAAATTTGAAAGTGGAAGTCT 210

QY 1134 atcctggaccatcgagagaggaagcaagcaccagaatctctcacagctacagacctca 1193

|||||

Db 211 ATCTGGAACCATGGAGGAGGACCAAGCAACCAAGAAATCTATCTACAGCTACAGACCTCA 270

QY 1194 aaagctgatcagcaaaacactagagggaagaacaatcttggatgtggggacaattcagt 1253

|||||

Db 271 AAAGGCTGATCAGCAAGCACTAGAGGAAGAACAAATCTTTGGATGTGGGGACAATTCAGT 330

QY 1254 tcaactgatg 1262

|||||

Db 331 TCACTGATG 339

RESULT 6

HSIMPG02 537 bp DNA PRI 28-OCT-1998

LOCUS Homo sapiens interphotoreceptor matrix gene (IPM150), exon 2.

DEFINITION AF017761

ACCESSION AF017761

VERSION NC_017761.1

KEYWORDS 2 of 17

SEGMENT human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 537)

AUTHORS Gehrig, A., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M., Schmid, M., and Weber, B.H.

TITLE Genomic organization and chromosomal localization of the interphotoreceptor matrix proteoglycan-1 (IMPG1) gene: a candidate for 6q-linked retinopathies

JOURNAL Cytogetnet. Cell Genet. 81 (1), 12-17 (1998)

MEDLINE 98358139

REFERENCE 2 (bases 1 to 537)

AUTHORS Gehrig, A., Feilbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussels, I.E. and Weber, B.H.F.

TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150) localized to 6q14.2-q15 in autosomal dominant Stargardt-like macular dystrophy, progressive bifocal choriorretinal atrophy (PBCRA), and North Carolina macular dystrophy (MCDRI) Unpublished

JOURNAL 3 (bases 1 to 537)

REFERENCE Feilbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.

AUTHORS Direct Submission

TITLE Submitted (09-AUG-1997) Humanenetik, Universitaet Wuerzburg, Am Hubland, Wuerzburg D-97074, Germany

JOURNAL Location/Qualifiers

FEATURES source

1. .537

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="6q14.2-q15"

97. .330

/gene="IPM150"

BASE COUNT 195 a 99 c 74 g 169 t

ORIGIN

Query Match 7.2%; Score 236; DB 11; Length 537;

Best Local Similarity 100.0%; Pred. No. 5.9e-119;

Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 agatatctcattaaacatataccattctgaacttaagacacacagatcccccaagaaa 252

|||||

Db 95 AGATATCTCCATTAAACATATACCATCTCTGAACCTAAAGACATAGACAAATCCCCAAGAAA 154

QY 253 tgaacaactgaaagtactgaaaaaatgtacaaaatgtcaaaatgtcaactatgagcgaatatccga 312

|||||

Db 155 TGAACAACCTGAAAGTACTGAAAAAATGTACAAAATGTCAAACTATGATGAGAGGAATATTCGA 214

QY 313 ttggcaagacatcgacaacaaagatccgcatttttcccaacgggggttaaaatctctcc 372

|||||

Db 215 TTTGGCAAGCATCGCAACAAAAGATCCGCATTTTTCCCAACGGGGTTAAAGTCTGCC 274

QY 373 acaggaatccatgaacacagatttttagacagcttctcaagcttattatagattgagag 428

|||||

Db 275 ACAGGAATCCATGAACAGATTTTAGACAGCTCTCAAGCTTATATAGATTGAGAG 330

RESULT 7

AL157379/c 194704 bp DNA HTG 22-MAR-2000

LOCUS Homo sapiens chromosome 6 clone RPI-62L18, *** SEQUENCING IN PROGRESS ***; 28 unordered pieces.

DEFINITION AL157379

ACCESSION AL157379.2

VERSION GI:7159486

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 194704)

AUTHORS Sims, S.

TITLE Direct Submission

JOURNAL Submitted (22-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT Requests: clonerequest@sanger.ac.uk

On Mar 6, 2000 this sequence version replaced gi:7009540.

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate

BASE COUNT 127 a 119 c 117 g 114 t

Query Match 4.7%; Score 153; DB 11; Length 477;
Best Local Similarity 100.0%; Pred. No. 6.5e-73;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2169 cagctgatcaagcagatccctgcgaagtcctggcctgcggcgaatttgcgaatgtgtaa 2228

Db 68 CAGCTGATCAGCAGATCCCTGCAAGTTCCTGGCTCGCGGGAATTTGCCCAATGTGTA 127

QY 2229 agaacgaacgactgaggaagcggagtgctgcctgcaaacaggatatacacagccaggga 2288

Db 128 AGACGAACGGACTGAGGAAGCGGAGTCTGCTGCAACACAGGATATGACAGCCAGGGA 187

QY 2289 gcttgagcgtctggaacagcgcctctgtgccc 2321

Db 188 GCCTGGAGGCTCTGGACACGAGCCTCTGTGGCC 220

RESULT 11

HSIMP07
LOCUS Homo sapiens 377 bp DNA PRI 28-OCT-1998
DEFINITION Homo sapiens interphotoreceptor matrix gene (IPM150), exon 7.
ACCESSION AF017766
VERSION AF017766.1 GI:3800721

KEYWORDS

SEGMENT 7 of 17

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 377)

Felbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,

Schmid, M. and Weber, B.H.

Genomic organization and chromosomal localization of the

interphotoreceptor matrix proteoglycan-1 (IMPG1) gene: a candidate

for 6q-linked retinopathies

Cytogenet. Cell Genet. 81 (1), 12-17 (1998)

98358139

2 (bases 1 to 377)

Gehrig, A., Felbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussells, I.E.

and Weber, B.H.F.

Assessment of a novel interphotoreceptor matrix gene (IPM150)

localized to 6q14.2-q15 in autosomal dominant Stargardt-like

macular dystrophy, progressive bifocal choriorretinal atrophy

(PBCRA), and North Carolina macular dystrophy (MCDRI)

Unpublished

3 (bases 1 to 377)

Felbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.

Direct Submission

Submitted (09-AUG-1997) HumanGenetik, Universitaet Wuerzburg, Am

Hubland, Wuerzburg D-97074, Germany

Location/Qualifiers

1. 377

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="6q14.2-q15"

163.303

/gene="IPM150"

BASE COUNT 110 a 86 c 83 g 98 t

exon

Query Match 4.3%; Score 141; DB 11; Length 377;

Best Local Similarity 100.0%; Pred. No. 3e-66;

Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 794 gaaagagaacagaattcgtctgttgaggagcagaggtgagctcagctctctg 853

Db 163 GAAAGAGAAACAGATTCTGCTGTGTGGAGGACAGAGGCTCAGCTCTCTCTG 222

QY 854 gtaaaccaagaagtccaagcagagctcgtgactccagtcaccagtcaccattaccaggagcta 913

Db 223 GTAAACCAAGAAGTCAAGCAGAGCTCGTCTGACTCCAGTCCCATATATACCAGGAGCTA 282

QY 914 gcaggaagtcaccaacttcag 934

Db 283 GCAGGAAGTCCCACTTCAG 303

RESULT 12

AF047491

LOCUS

DEFINITION

Macaca fascicularis interphotoreceptor matrix proteoglycan-150

ACCESSION

AF047491

VERSION

AF047491.1

KEYWORDS

SOURCE

ORGANISM

crab-eating macaque.

Macaca fascicularis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;

Macaca.

REFERENCE

1 (bases 1 to 555)

Kuehn, M.H. and Hageman, G.S.

Characterization And Complete cDNA Sequence Of IPM 150, A Novel

Human Photoreceptor Cell-Associated Chondroitin-Sulfate

Proteoglycan

Unpublished

REFERENCE

2 (bases 1 to 555)

Kuehn, M.H. and Hageman, G.S.

Direct Submission

Submitted (10-FEB-1998) Ophthalmology and Visual Sciences,

University of Iowa, 200 Hawkins Drive, 1110E PPP, Iowa City, IA

52240, USA

Location/Qualifiers

1. .555

/organism="Macaca fascicularis"

/db_xref="taxon:9541"

/tissue type="retina"

<1. .>555

/note="IPM 150"

/codon_start=1

/product="interphotoreceptor matrix proteoglycan 150"

/protein_id="AAC03788.1"

/db_xref="GI:2906320"

/translation="IFFPVGKVCPOESMKIILASIQAYRLRVCOEAVWEAYRFLD

RIPDTGEYODWYSCQOETFLFDIGONFSNQHLLDLOQRKQSFPERKDEVTE

KTLGEPSETIVYSTDVAVSLGLFPFVPTDDTLNLLNDLNALNDKMPPTTERETELAVS

EEQVVELSISLINQRFKAEADSQS"

BASE COUNT 162 a 131 c 131 g 131 t

ORIGIN

Query Match

3.4%; Score 112; DB 11; Length 555;

Best Local Similarity 100.0%; Pred. No. 3.6e-50;

Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 400 cagcttcacgcttatatagattgagtgctgcaggaagcagtgatgcgaagcatatcg 459

Db 60 CAGCTCTCAAGCTTATTATAGATTGAGTGTCTCAGGAAGCAGTATGGAAGCATATCG 119

QY 460 gatctttctggtatcgatccctcacacaggggaatatcagactgggtcagc 511

Db 120 GATCTTTCTGGATCGCATCCCTGACACAGGGGAATATCAGGACTGGGTGACG 171

RESULT 13

HSIMP06

LOCUS

DEFINITION

Homo sapiens interphotoreceptor matrix gene (IPM150), exon 6.

ACCESSION

AF017765

VERSION

AF017765.1 GI:3800720

KEYWORDS
SEGMENT
SOURCE

6 of 17
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 311)
Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
Schmid,M. and Weber,B.H.

TITLE

Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPg1) gene: a candidate
for 6q-linked retinopathies

JOURNAL
MEDLINE

Cytogenet. Cell Genet. 81 (1), 12-17 (1998)

REFERENCE
AUTHORS

2 (bases 1 to 311)
Gehrig,A., Felbor,U., Kelsell,R., Hunt,D.M., Maunenee-Hussells,I.E.
and Weber,B.H.F.

TITLE

Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRL)

JOURNAL
REFERENCE

3 (bases 1 to 311)
Felbor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.

TITLE

Direct Submission
Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany

JOURNAL

Hubland, Location/Qualifiers

FEATURES
source

1..311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
139..242
/gene="IPM150"
95 a 70 c 45 g 101 t

exon

BASE COUNT 95 a 70 c 45 g 101 t
ORIGIN

Query Match

Best Local Similarity 3.3%; Score 109; DB 11; Length 311;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

686 acagatgttgcacacgtctcaacttgccttccctcactcctgacacccctc 745

Db

135 ACAGATGTTCGACACGTCTCACTTGGCCCTTCCTCTCTCTCTGATGACCCCTCCTC 194

QY

746 aatgaattctcgataatacactcaacgacacacacgaagatgcctacaacag 794

Db

195 AATGAATTCGATAATACTCACTCAACGACACCAAGATGCTACACAG 243

RESULT 14

HSIMPG12

LOCUS HSIMPG12 438 bp DNA PRI 28-OCT-1998

DEFINITION

Homo sapiens interphotoreceptor matrix gene (IPM150), exon 12.

ACCESSION

AF017771

VERSION

AF017771.1 GI:3800726

KEYWORDS

12 of 17

SEGMENT

human.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE

1 (bases 1 to 438)

AUTHORS

Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
Schmid,M. and Weber,B.H.

TITLE

Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPg1) gene: a candidate
for 6q-linked retinopathies

JOURNAL

Cytogenet. Cell Genet. 81 (1), 12-17 (1998)

MEDLINE

98358139

REFERENCE

2 (bases 1 to 438)

AUTHORS

Gehrig,A., Felbor,U., Kelsell,R., Hunt,D.M., Maunenee-Hussells,I.E.
and Weber,B.H.F.

TITLE

Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRL)

JOURNAL

Unpublished

REFERENCE

3 (bases 1 to 438)

AUTHORS

Felbor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.

TITLE

Direct Submission

JOURNAL

Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany

FEATURES

Location/Qualifiers

source

1..438

exon

BASE COUNT 140 a 92 c 69 g 136 t 1 others
ORIGIN

Query Match

Best Local Similarity 2.5%; Score 81; DB 11; Length 438;

Matches

81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1338 aggatgtctatttgatccagaacttctctctgttgaaacccagcttgagacagtggacg 1397

Db

269 AGGATGCTACTTTGAGTCCAGACTTCTCTCTGTGTGAACCCAGCTTGAGACAGTGGACG 328

QY

1398 gacgagacatgtctacctg 1418

Db

329 GAGCAGACATGGTCTACCTG 349

RESULT 15

HSIMPG11

LOCUS

HSIMPG11 347 bp DNA PRI 28-OCT-1998

DEFINITION

Homo sapiens interphotoreceptor matrix gene (IPM150), exon 12.

ACCESSION

AF017770

VERSION

AF017770.1 GI:3800725

KEYWORDS

11 of 17

SEGMENT

human.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 347)

AUTHORS

Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
Schmid,M. and Weber,B.H.

TITLE

Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPg1) gene: a candidate
for 6q-linked retinopathies

JOURNAL

Cytogenet. Cell Genet. 81 (1), 12-17 (1998)

MEDLINE

98358139

REFERENCE

2 (bases 1 to 347)

AUTHORS

Gehrig,A., Felbor,U., Kelsell,R., Hunt,D.M., Maunenee-Hussells,I.E.
and Weber,B.H.F.

TITLE

Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRL)

JOURNAL

Unpublished

REFERENCE

3 (bases 1 to 347)

AUTHORS

Felbor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.

TITLE

Direct Submission
Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany

FEATURES

Location/Qualifiers

source

1..347

exon

BASE COUNT 140 a 92 c 69 g 136 t 1 others
ORIGIN

Query Match

Best Local Similarity 2.5%; Score 81; DB 11; Length 438;

Matches

81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1338 aggatgtctatttgatccagaacttctctctgttgaaacccagcttgagacagtggacg 1397

Db

269 AGGATGCTACTTTGAGTCCAGACTTCTCTCTGTGTGAACCCAGCTTGAGACAGTGGACG 328

QY

1398 gacgagacatgtctacctg 1418

Db

329 GAGCAGACATGGTCTACCTG 349

```
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
148..224
/feature="IPW150"
BASE COUNT 104 a 78 c 51 g 112 t 2 others
ORIGIN

Query Match 2.4%; Score 79; DB 11; Length 347;
Best Local Similarity 100.0%; Pred. No. 7.3e-32;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1262 gaattgctggtacactgccagcctttggtcctgacaccccaatcagagctgcccacatct 1321
|||||
DB 147 GAAATTGCTGGATCAGTCCAGCCTTTGGTCTCTGACACCCCAATCAGAGCTGCCACATCT 206
|||||

QY 1322 ttgctgttataacagagg 1340
|||||
DB 207 TTTGCTGTTTATACAGAGG 225
|||||
```

Search completed: September 3, 2000, 16:29:14
Job time: 21871 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 14:06:08 ; Search time 196.21 Seconds
(without alignments)
707.694 Million cell updates/sec

Title: US-09-183-972-1
Perfect score: 555
Sequence: 1 attttcttcgaacggggt.....agctcgtagctctcagta 555

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18	3.2	2754	1 T58505	H218 cDNA encoding
C 2	18	3.2	2754	1 V84039	DNA encoding rat H
C 3	17	3.1	158	1 T75262	Nucleotide sequenc
C 4	17	3.1	355	1 T19541	Human gene signatu
C 5	17	3.1	430	1 V86891	EST clone BI41. Ne
C 6	17	3.1	722	1 T15230	Tumor necrosis fac
C 7	17	3.1	722	1 T94633	TNF-R1-DD ligand p
C 8	17	3.1	1023	1 Q28271	Encodes human IGFB
C 9	17	3.1	1023	1 T15231	Insulin-like growt
C 10	17	3.1	1023	1 T94634	TNF-R1-DD ligand p
C 11	17	3.1	1611	1 Q26814	IGFBP6 DNA. Insuli
C 12	17	3.1	1650	1 Q65219	Human insulin-like
C 13	17	3.1	5144	1 Q94446	Bacterial transfer
C 14	17	3.1	5144	1 T49504	Transferrin recept
C 15	17	3.1	5144	1 V21440	H. influenzae stra
C 16	17	3.1	5893	1 V59979	Nucleic acid encod
C 17	17	3.1	5893	1 T75237	Nucleotide sequenc
C 18	17	3.1	7785	1 V81446	Pig p105 zona pell
C 19	16	2.9	294	1 X40261	Human secreted pro
C 20	16	2.9	406	1 X20757	Polynucleotide seq
C 21	16	2.9	428	1 X41104	Human secreted pro
C 22	16	2.9	505	1 V90020	EST clone CW1381.
C 23	16	2.9	572	1 Q44852	Bacillus firmus so
C 24	16	2.9	587	1 X20420	Human secreted pro
C 25	16	2.9	598	1 V60575	Human proline-rich
C 26	16	2.9	1113	1 Q62305	Clostridium perfr
C 27	16	2.9	1206	1 V03793	PAMV coat protein
C 28	16	2.9	1272	1 Q66344	G3 pGRF receptor g
C 29	16	2.9	1290	1 Q66343	pGRF receptor gene
C 30	16	2.9	1356	1 Q66342	pGRF receptor gene
C 31	16	2.9	1425	1 T35220	Cytoplasmic antipr
C 32	16	2.9	1545	1 V81449	Ovine growth hormo
C 33	16	2.9	1630	1 Q28270	Encodes rat IGFBP-

C 34	16	2.9	1912	1 X19251	cDNA sequence SEQ
C 35	16	2.9	1996	1 T90830	cDNA encoding gluc
C 36	16	2.9	2020	1 X00649	Human secreted pro
C 37	16	2.9	2184	1 X08909	Human FADD-Interac
C 38	16	2.9	2194	1 X19248	Human EAAT5 glutam
C 39	16	2.9	2597	1 T79135	Bovine male enhanc
C 40	16	2.9	2745	1 T90831	Glucosylase P gen
C 41	16	2.9	3260	1 Q79325	Mammalian MEK kina
C 42	16	2.9	3260	1 T05570	MEKK1 cDNA. New si
C 43	16	2.9	3260	1 V22676	cDNA encoding a mu
C 44	16	2.9	3260	1 V45619	MEKK1 protein codi
C 45	16	2.9	3411	1 V16518	DNA encoding a Bac

ALIGNMENTS

RESULT 1

T58505/C

ID T58505 standard; DNA; 2754 BP.

AC T58505;

DT 01-APR-1997 (first entry)

DE H218 cDNA encoding p(H218), G-protein coupled receptor.

KW p(H218); G-protein coupled receptor; cell differentiation; proliferation;

KW Proline directed kinase; cell division; growth factor response; rat-edg;

KW therapy; diagnosis; ss.

OS Rattus rattus.

FH Key

FT 5'utr

FT Location/Qualifiers

FT 1..147

FT /*tag= a

FT /*tag= 143..151

FT /*tag= b

FT /note= "optimal consensus sequence for translation

FT initiation"

FT 148..1206

FT cds

FT /*tag= c

FT /product= p(H218)

FT 1207..2754

FT 3'utr

FT /*tag= d

FT repeat_unit

FT 868..872

FT /*tag= e

FT repeat_region

FT 868..2083

FT /*tag= f

FT repeat_region

FT 2207..2248

FT /*tag= g

FT misc_feature

FT 2393..2397

FT /*tag= h

FT /note= "consensus sequence associated with mRNA

FT instability"

FT 2724..2729

FT polya_signal

FT /*tag= i

US5585476-A.

PD 17-DEC-1996.

PF 15-FEB-1994; 196989.

PR 15-FEB-1994; US-196989.

PA (MACL/) MACLENNAN A J.

PI Maclellenn AJ;

DR WPI: 97-051235/05.

DR P-PSDB; W01663.

DR DNA encoding rat protein p(H218) - associated with cell

PT proliferation and/or differentiation

PS Example 1; Column 15-18; 33pp; English.

CC This sequence is a novel rat cDNA H218 which encodes p(H218), a member

CC of the G-protein coupled receptor superfamily of proteins. The amino acid

CC similarity between p(H218) (W01663) and p(rat-edg) (W01664) suggests that

CC they may be activated by the same endogenous ligand(s). The expression

CC pattern of mRNA transcripts of both genes in cell lines, various rat

CC tissues and developing rat brain suggests that they both play a role in

CC cell proliferation and/or differentiation.

CC Sequence 2754 BP; 569 A; 794 C; 714 G; 677 T;

SQ

Query Match

3.2%; Score 18; DB 1; Length 2754;

Best Local Similarity 100.0%; Pred. No. 7.4; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0;

QY 217 ttccagcaattcccaggag 234
|||||
Db 2560 TTCAGCAATTCCCAGGAG 2543

RESULT 2

V84039/c
ID V84039 standard; DNA; 2754 BP.
AC V84039
DT 12-MAR-1999 (first entry)
DE DNA encoding rat H218, G-protein coupled receptor superfamily member.
KW H218; G-protein coupled receptor superfamily; cell proliferation;
KW cell differentiation; cancer; diagnosis; therapy; ss.
OS Rattus sp.
FH Key Location/Qualifiers
FT CDS 148..1206
FT FT /*tag= a
FT FT /product= H218
FT FT 2393..2397
FT FT /*tag= b
FT FT /note= "consensus sequence associated with mRNA
FT polyA_signal 2724..2729
FT FT /*tag= c
FT FT
FT FT
PN US856443-A.
PD 05-JAN-1999.
PF 06-DEC-1996; 760936.
PR 15-FEB-1994; US-196989.
PR 06-DEC-1996; US-760936.
PA (MACL/) MACLENNAN A J.
PI MacLennan AJ.
DR WPI; 99-105192/09.
DR P-PSDB; W87790.
PT New isolated polynucleotide encoding a G-protein coupled receptor -
PT used to develop products for use in diagnosis and therapy of
PT conditions involving cell proliferation and differentiation
PS Claim 1; Columns 15-18; 33pp; English.
CC The present sequence encodes a protein designated H218. The H218
CC polypeptide is a member of the G-protein coupled receptor superfamily.
CC The H218 polypeptide is involved in cell proliferation and
CC differentiation, and in disease states such as cancer. The
CC polynucleotides and polypeptides can be used to develop products
CC for diagnosis and therapy.
SQ Sequence 2754 BP; 569 A; 794 C; 714 G; 677 T;

Query Match 3.2%; Score 18; DB 1; Length 2754;

Best Local Similarity 100.0%; Pred. No. 7.4; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0;

QY 217 ttccagcaattcccaggag 234
|||||
Db 2560 TTCAGCAATTCCCAGGAG 2543

RESULT 3

T75262/c
ID T75262 standard; DNA; 158 BP.
AC T75262;
DT 21-DEC-1998 (first entry)
DE Nucleotide sequence encoding human RAD50 exon 11.
KW ds; human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;
KW central nervous system.
OS Homo sapiens.
PN WQ9727284-A2.
PD 31-JUL-1997.
PF 24-JAN-1997; U01299.
PR 17-JUL-1996; US-687080.
PR 26-JAN-1996; US-592126.
PA (GENE-) GENELABS TECHNOLOGIES INC.

PI Dolganov G;
DR WPI; 97-393672/36.
PT Human tumour suppressor gene RAD50 - useful to detect
PT predisposition to, decrease risk of and treat cancer, also Septin-2
PT homologues
PS Claim 1; Page 112; 195pp; English.
CC The human RAD50 (hRAD50) is involved in DNA repair and has tumour
CC suppression activity, can be used to detect predisposition to, decrease
CC the risk of or treat cancers, e.g. acute myeloid leukaemia,
CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,
CC therapy related acute myeloid leukaemia, refractory anaemia or refractory
CC anaemia with excess blasts. Also disclosed in this invention is Human
CC Septin-2 homologues of which may be used as targets for cancer therapies
CC and central nervous system directed treatment methods, and to measure the
CC proliferative potential of selected cell types.
SQ Sequence 158 BP; 69 A; 28 C; 28 G; 33 T;

Query Match 3.1%; Score 17; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 tagccagcttcacagct 72
|||||
Db 105 TAGCCAGCTTCAAGCT 89

RESULT 4

TI9541/c
ID TI9541 standard; cDNA to mRNA; 355 BP.
AC TI9541;
DT 28-JUN-1996 (first entry)
DE Human gene signature HUMG00604.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 412; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in TI9001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 355 BP; 109 A; 60 C; 64 G; 118 T;

Query Match 3.1%; Score 17; DB 1; Length 355;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 aatccatgaacagatt 54

Db 85 AATCCATGAACAGATT 69
 |||

RESULT 5

V86891/c
 ID V86891 standard; cDNA; 430 BP.
 AC V86891;
 DT 27-APR-1999 (first entry)
 DE EST clone B141.
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; ESN; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS Homo sapiens.
 PN WO9845435-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06954.
 PR 10-APR-1998; US-835913.
 PA (GEMV) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 WPI; 99-070076/06.
 DR New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 382; 633pp; English.
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cacharin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.
 SQ Sequence 430 BP; 117 A; 124 C; 111 G; 77 T;

Query Match 3.1%; Score 17; DB 1; Length 430;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 gccttccctgtcactc 388

Db 194 GCCTTCCCTGTCATC 178

RESULT 6

T15230/c
 ID T15230 standard; cDNA; 722 BP.
 AC T15230;
 DT 09-DEC-1996 (first entry)
 DE Tumor necrosis factor receptor 1 death domain ligand cDNA (clone 20DD).
 KW TNF-R1-DD; tumor necrosis factor receptor 1 death domain; inhibitor;
 KW P55; anti-inflammatory; autoimmune disease; graft versus host reaction;
 KW osteoporosis; cachexia; diabetes; sequence identity; IGFBP-5;
 KW insulin-like growth factor binding protein-5; ds.
 OS Homo sapiens.
 PN WO9612735-A1.
 PD 02-MAY-1996.
 PF 12-OCT-1995; U12724.
 PR 19-OCT-1994; US-327514.
 PR 19-JUN-1995; US-494440.
 PR 26-SEP-1995; US-533501.

Query Match 3.1%; Score 17; DB 1; Length 430;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 gccttccctgtcactc 388
 Db 194 GCCTTCCCTGTCATC 178

PA (GEMV) GENETICS INST INC.
 PI Chen J, Graham J, Lin L, Schievella AR;
 DR WPI; 96-230551/23.
 DR P-PSDB; R95328.
 PT TNF receptor death domain ligand proteins and inhibitors of ligand
 PT binding - for prevention and treatment of pref. anti-inflammatory
 PT conditions, e.g. auto-immune disease, graft versus host reaction
 PT osteoporosis, etc.
 PS Claim 27; Page 39-40; 83pp; English.
 CC The present sequence, clone 20DD (ATCC 69704), encodes a tumour necrosis
 CC factor (TNF) receptor 1 (R1) death domain (DD) ligand (see R95328). A
 CC yeast genetic selection method, the "interaction trap", was used to
 CC screen W138 cell cDNA libraries for proteins that interact with the DD
 CC of the P55 type 1 TNF-R. The protein (R95328) encoded by clone 20DD is
 CC identical to amino acids 87-272 of insulin-like growth factor binding
 CC protein-5 (IGFBP-5; R95329). Based upon the sequence identity between
 CC clone 20DD and IGFBP-5, IGFBP-5 and certain fragments of it will exhibit
 CC TNF-R1-DD ligand binding activity. TNF-R1-DD ligands and their inhibitors
 CC are useful in the prevention and treatment of anti-inflammatory
 CC conditions and other conditions such as cachexia, autoimmune disease,
 CC graft versus host reaction, osteoporosis, diabetes, etc.
 SQ Sequence 722 BP; 197 A; 228 C; 182 G; 115 T;

Query Match 3.1%; Score 17; DB 1; Length 722;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 ggaactgggtcagcttct 175

Db 232 GGACGGGTCAGCTTCT 216

RESULT 7

T94633/c
 ID T94633 standard; cDNA; 722 BP.
 AC T94633;
 DT 19-MAR-1998 (first entry)
 DE TNF-R1-DD ligand protein clone 20DD coding sequence.
 KW Tumor necrosis factor receptor P55 type; TNF-R1-DD ligand protein;
 KW death domain; TNF-R1; inhibitor identification; TNF-induced condition;
 KW insulin-like growth factor binding protein-5; inflammatory condition;
 KW IGFBP-5; therapy; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 2..562
 FT /*tag= a

PN WO9730084-A1.

PD 21-AUG-1997.

PF 11-FEB-1997; U02146.

PR 15-AUG-1996; US-698551.

PR 15-FEB-1996; US-602228.

PA (GEMV) GENETICS INST INC.

PI Chen J, Graham J, Lin L, Schievella AR;

DR WPI; 97-424976/39.

PT Tumor necrosis factor receptor P55 type death domain ligand

PT proteins - useful for preventing or ameliorating inflammatory

PT conditions

PS Claim 27; Page 42-43; 103pp; English.

CC This sequence represents the coding sequence of the invention. This
 CC sequence was isolated from clone 20DD and encodes a tumour necrosis
 CC factor receptor P55 type (TNF-R1) death domain (DD) ligand protein. A
 CC host cell containing this sequence is used for the recombinant production
 CC of TNF-R1-DD. The TNF-R1-DD ligand protein can be used in a method to
 CC identify inhibitors of TNF-R DD binding. The TNF-R1-DD ligand protein,
 CC insulin-like growth factor binding protein-5 (IGFBP-5), has TNF-R1-DD
 CC ligand activity) or inhibitors of TNF-R1-DD ligand protein are capable of
 CC preventing or ameliorating an inflammatory condition, preferably by
 CC inhibiting TNF-R DD binding. Identification and isolation of ligands
 CC allows their effects upon TNF-R signal transduction and use as
 CC therapeutic agents for treatment of TNF-induced conditions to be
 CC examined.

SQ Sequence 722 BP; 197 A; 228 C; 182 G; 115 T;

```

Query Match          3.1%; Score 17; DB 1; Length 722;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 ggactgggtcagcttct 175
    |||||
Db 232 GGACTGGGTCAGCTTCT 216

RESULT 8
Q28271/c
ID Q28271 standard; DNA; 1023 BP.
AC Q28271;
DE 16-FEB-1993 (first entry)
KW Encodes human IGFBP-5
KW human insulin-like growth factor binding protein-5; IGF-I; IGF-II;
KW breast cancer; bone cancer; modulating bone growth; purification;
KW affinity columns; antibodies; diagnosis; testing; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT cds 57..872
   /*tag= a
   WO9214834-A.
   PD 03-SEP-1992.
   PF 13-FEB-1992; U01196.
   PR 14-FEB-1991; US-658410.
   PA (WHIT-) WHITTIER INST DIABETES & ENDOCRINOLOGY.
   PI Ling NC, Shimaski S;
   DR P-PSDB; R26995.
   DT DNA encoding insulin-like growth factor binding protein - useful
   PT for treating breast and bone cancer and modulating bone growth
   PS Claim 3; Page 11; 4pp; English.
   CC This sequence encodes human insulin-like growth factor binding protein.
   CC The insert from Q28270 was excised and subsequently used as a probe
   CC to screen a human placenta cDNA library. Positive clones were prepared,
   CC purified, and one of the longest clones subcloned into pBluescript SK+
   CC for sequencing.
   SQ Sequence 1023 BP; 224 A; 352 C; 284 G; 163 T;

Query Match          3.1%; Score 17; DB 1; Length 1023;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 ggactgggtcagcttct 175
    |||||
Db 545 GGACTGGGTCAGCTTCT 529

RESULT 10
T94634/c
ID T94634 standard; cDNA; 1023 BP.
AC T94634;
DE 19-MAR-1998 (first entry)
DE TNF-R1-DD ligand protein clone IGFBP-5 coding sequence.
KW Tumour necrosis factor receptor p55 type; TNF-R1-DD ligand protein;
KW death domain; TNF-R1; inhibitor identification; TNF-induced condition;
KW insulin-like growth factor binding protein-5; inflammatory condition;
KW IGFBP-5; therapy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 57..875
   /*tag= a
   WO9730084-A1.
   PD 21-AUG-1997.
   PF 11-FEB-1997; U02146.
   PR 15-AUG-1996; US-698551.
   PR 15-FEB-1996; US-602228.
   PA (GEMY ) GENETICS INST INC.
   PI Chen J, Graham J, Lin L, Schievella AR;
   DR P-PSDB; W35572.
   DT Tumour necrosis factor receptor p55 type death domain ligand
   PT proteins - useful for preventing or ameliorating inflammatory
   PT conditions
   PS Claim 27; Page 44-45; 103pp; English.
   CC This sequence represents the coding sequence of the invention. This
   CC sequence encodes the insulin-like growth factor binding protein-5
   CC (IGFBP-5) which is a tumour necrosis factor receptor p55 type (TNF-R1)
   CC death domain (DD) ligand protein. A host cell containing this sequence is
   CC used for the recombinant production of TNF-R1-DD. The TNF-R1-DD ligand
   CC protein can be used in a method to identify inhibitors of TNF-R DD
   CC binding. The TNF-R1-DD ligand protein, IGFBP-5 (has TNF-R1-DD ligand
   CC activity), or inhibitors of TNF-R1-DD ligand protein are capable of
   CC preventing or ameliorating an inflammatory condition, preferably by
   CC inhibiting TNF-R DD binding. Identification and isolation of ligands
   CC allows their effects upon TNF-R signal transduction and use as
   CC therapeutic agents for treatment of TNF-induced conditions to be
   CC examined.
   SQ Sequence 1023 BP; 224 A; 352 C; 284 G; 163 T;

```



```

Query Match          3.1%; Score 17; DB 1; Length 1023;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 ggactgggtcagcttct 175
DB 545 GGACTGGGTGAGCTTCT 529

RESULT 11
Q26814/c
ID Q26814 standard; DNA; 1611 BP.
AC Q26814;
DT 20-JAN-1993 (first entry)
DE IGFBP6 DNA.
KW Insulin; like; growth factor; binding protein; BP-6; IGF; growth;
KW regeneration; hypopituitarism; osteoporosis; anaemia; breast cancer;
KW kidney cancer; diabetic retinopathy; purification; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT cds 44..862
FT /*tag= a
FT PN WO212243-A.
FT PD 23-JUL-1992.
FT PF 02-JAN-1992; U00107.
FT PR 08-JAN-1991; US-638628.
FT PA (CHIR ) CHIRON CORP.
FT PI Kiefer MC;
FT DR WPI; 92-268666/32.
FT DR N-PSDB; R25700.
FT PT Insulin-like growth factor binding protein-6 - for treating
FT PT hypopituitarism, osteoporosis, anaemia(s), cancer, etc.,
FT PT stimulating growth and wound healing, also useful in diagnosis
FT PS Claim 8; Fig 1; 65pp; English.
FT CC PCR was used to isolate insulin-like growth factor binding protein
FT CC 6 (IGFBP6) from a human osteosarcoma cDNA template using PCR
FT CC primers whose sequences were based on regions of high homology
FT CC between the amino acid sequences of five known IGFBPs.
FT CC Two probes were synthesised, based on the sequence of the PCR prod.
FT CC and were used to screen a ZAPII/human osteosarcoma cDNA library.
FT CC Twelve double positive clones were found and were analysed by
FT CC restriction digestion. A unique sequence of ca. 1.7 kb designated
FT CC IGFBP6 was identified. IGFBP6 may be used alone, or with IGF to
FT CC stimulate growth, tissue or organ regeneration or wound healing.
FT CC Also IGFBP6 has applications in the treatment and diagnosis of
FT CC hypopituitarism, osteoporosis, anaemias, and disorders due to
FT CC excessive prodn. of free IGF, e.g. breast or kidney cancer,
FT CC diabetic retinopathy, and abnormal growth of tall subjects.
FT CC The binding protein can also be used to purify IGF e.g. by
FT CC affinity chromatography. See also Q26810-4.
FT SQ Sequence 1611 BP; 408 A; 467 C; 454 G; 282 T;

Query Match          3.1%; Score 17; DB 1; Length 1611;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 ggactgggtcagcttct 175
DB 532 GGACTGGGTGAGCTTCT 516

RESULT 12
Q65519/c
ID Q65519 standard; DNA; 1650 BP.
AC Q65519;
DT 28-NOV-1994 (first entry)
DE Human insulin-like growth factor binding protein-5 DNA.
KW Insulin-like growth factor binding protein-5; IGFBP-5; hormone; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT mat_peptide 549..132

```

```

FT /*tag= a
FT /note= "preferred truncated IGFBP-5, claim 12"
FT 123..627
FT /*tag= b
FT /note= "preferred truncated IGFBP-5, claim 13"
FT 63..881
FT /*tag= c
FT PN WO9410207-A.
FT PD 11-MAY-1994.
FT PF 29-OCT-1993; U10462.
FT PR 04-NOV-1992; US-972142.
FT PA (CHIR ) CHIRON CORP.
FT PI Address DL; Kiefer MC;
FT DR WPI; 94-167395/20.
FT DR P-PSDB; R55084.
FT PT Truncated insulin-like growth factor binding protein - has
FT PT reduced affinity for insulin-like growth factor, useful for
FT PT stimulating bone cell growth and mitogenic activity
FT PS Disclosure; Fig. 1; 56pp; English.
FT CC This sequence is contained in plasmid pBSBP6-1A and encodes 2
FT CC preferred truncated IGFBP-5 proteins derived from human U-2
FT CC osteosarcoma cells. This sequence may be optionally fused to a
FT CC calcitonin gene for expression in a recombinant host, preferably
FT CC yeast or CHO.
FT SQ Sequence 1650 BP; 419 A; 476 C; 463 G; 292 T;

Query Match          3.1%; Score 17; DB 1; Length 1650;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 ggactgggtcagcttct 175
DB 551 GGACTGGGTGAGCTTCT 535

RESULT 13
Q94446
ID Q94446 standard; DNA; 5144 BP.
AC Q94446;
DT 09-OCT-1996 (first entry)
DE Bacterial transferrin receptor operon (H. influenzae strain SB33).
KW Tbp1; Tbp2; transferrin receptor operon; vaccine; antigen;
KW non-typable strain; Haemophilus influenzae; meningitis; ss.
OS Haemophilus influenzae strain SB33.
FH Key Location/Qualifiers
FT cds 192..698
FT /*tag= a
FT /product= Tbp2
FT 2135..4870
FT /*tag= b
FT /product= Tbp1
FT PN WO9513370-A1.
FT PD 18-MAY-1995.
FT PF 07-NOV-1994; CA0616.
FT PR 08-NOV-1993; US-148968.
FT PR 29-DEC-1993; US-175116.
FT PA (CONN-) CONNAUGHT LAB LTD.
FT PI Chong P, Gray-owen S, Harkness R, Klein M, Loosmore S;
FT PI Murdin A, Schryvers A, Yang Y;
FT DR WPI; 95-194089/25.
FT DR P-PSDB; R77891-92.
FT PT Nucleic acids encoding Haemophilus transferrin receptor - used to
FT PT develop prods for detection and in diagnosis, prevention and
FT PT treatment of Haemophilus infection.
FT PS Claim 12; Fig 7A-N; 231pp; English.
FT CC The present DNA shows the transferrin receptor (Tfr) operon consisting of
FT CC two genes (Tbp1 and Tbp2) arranged in tandem and which are transcribed
FT CC from a single promoter, from the non-typable Haemophilus influenzae
FT CC strain SB33. The SB33 tbp2 gene has a single base deletion which resulted
FT CC in a frame-shift at residue 126 and premature truncation of the resulting
FT CC protein at residue 168. H. influenzae Tfr is iron- and/or haemin-
FT CC regulated and a putative fur-binding site has been identified upstream of

```

CC tbp2. Antibodies blocking this binding site may prevent bacterial growth.
 CC Fragments of the tfr (or its genes) are useful in vaccines to provide
 CC protection against, e.g. bacterial meningitis. An advantage of using the
 CC tfr is that it shares homology with tfr of other H. influenzae strains
 CC including non-typable strains.
 SQ Sequence 5144 BP; 1813 A; 888 C; 973 G; 1470 T;

Query Match 3.1%; Score 17; DB 1; Length 5144;

Best Local Similarity 100.0%; Pred. No. 24;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 gaaagagaaacagaact 470

|||||

DB 929 GAAAGAGAAACAGAACT 945

RESULT 14

T49504

ID T49504 standard; DNA; 5144 BP.

AC T49504;

DT 05-MAY-1998 (first entry)

DE Transferrin receptor gene sequence encoding Tbp1 and Tbp2.

KW Transferrin receptor; Haemophilus influenzae type b;

KW iron; human transferrin; iron source; antibody; bacterial growth;

KW vaccine; immunogenic truncated analogue; antigen; Tbp1; Tbp2; ss..

OS Haemophilus influenzae.

FH Key Location/Qualifiers

FT CDS 192..698

FT /*tag= a

FT /note= "encodes Tbp2 (W08966)"

FT CDS 2135..4870

FT /*tag= b

FT /note= "encodes Tbp1 (W08965)"

PN W09640929-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; CA0399.

PR 17-MAY-1996; US-649518.

PR 07-JUN-1995; US-483577.

PA (CONN-) CONNAUGHT LAB LTD.

PI Chong P, Gray-Owen S, Harkness RE, Klein MH, Loosmore SM,

PI Murdin AD, Schryvers AB, Yang Y;

DR P-PSDB; W08965-66.

DR Haemophilus truncated transferrin receptor protein analogue, Tbp2

PT used to induce protection against disease caused by transferrin

PT producing pathogens, or as antigen to detect Haemophilus tfr

PT antibodies

PS Claim 5; Fig 7A-N; 228pp; English.

CC The present sequence represents the transferrin receptor gene of

CC Haemophilus influenzae type b, strain SB33. The bacterial transferrin

CC receptor is composed of 2 chains, Tbp1 and Tbp2. H. influenzae is a

CC non-encapsulated or non-typable bacterium responsible for a wide range

CC of human diseases. Iron is an essential nutrient for the growth of these

CC bacteria, and they can utilise human transferrin as a source of iron.

CC Antibodies which block the access of the transferrin receptor to

CC its iron source prevent bacterial growth. The transferrin receptor, or

CC fragments, therefore, are good vaccine candidates. The full length Tbp2

CC protein is produced in low amounts in Escherichia coli. However, the

CC yield can be enhanced by truncation of the 3' end of the gene. An

CC immunogenic composition comprising (or encoding) the immunogenic

CC truncated analogue can be used to induce protection against a disease

CC caused by a bacterial pathogen that produces the transferrin receptor.

CC The immunogenic truncated analogue is also useful as an antigen in

CC immunoassays for the detection of Haemophilus transferrin receptor

CC antibodies, while the nucleic acid molecule can be used as a

CC hybridisation probe for the detection of other transferrin receptor

CC genes.

SQ Sequence 5144 BP; 1813 A; 888 C; 974 G; 1469 T;

Query Match

Best Local Similarity 3.1%; Score 17; DB 1; Length 5144;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 gaaagagaaacagaact 470
 |||||

DB 929 GAAAGAGAAACAGAACT 945

RESULT 15

V21440

ID V21440 standard; cDNA; 5144 BP.

AC V21440;

DT 20-JUL-1998 (first entry)

DE H. influenzae strain SB33 transferrin receptor operon.

KW tbp1; tbp2; vaccine; H. influenzae; antibody; diagnosis;

KW passive immunisation; ds.

OS Haemophilus influenzae.

FH Key Location/Qualifiers

FT CDS 192..698

FT /*tag= a

FT /product= Tbp1

FT CDS 2135..4870

FT /*tag= b

FT /product= Tbp2

US5708149-A.

PN 13-JAN-1998.

PD 07-JUN-1995; 487890.

PF 08-NOV-1994; US-337483.

PR 08-NOV-1993; US-148968.

PR 29-DEC-1993; US-175116.

PR 07-JUN-1995; US-487890.

PA (CONN-) CONNAUGHT LAB LTD.

PI Chong P, Gray-Owen S, Harkness R, Klein M, Loosmore S,

PI Murdin A, Schryvers A, Yang Y;

DR WPI; 98-100410/09.

DR P-PSDB; W54123, W54124.

DR Purification of recombinant Haemophilus transferrin-binding protein

PT - by solubilising inclusion bodies separated from cell lysate

PS Example 4; Fig 7; 261pp; English.

CC The H. influenzae transferrin receptor operon contains two genes (tbp1

CC and tbp2) under the transcriptional regulation of one promoter. The

CC proteins encoded from these genes can be expressed in a recombinant

CC host. The proteins can be used in vaccines against H. influenzae

CC infections or to produce antibodies for use in diagnosis or passive

CC immunisation.

SQ Sequence 5144 BP; 1813 A; 888 C; 974 G; 1469 T;

Query Match

Best Local Similarity 3.1%; Score 17; DB 1; Length 5144;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 gaaagagaaacagaact 470

|||||

DB 929 GAAAGAGAAACAGAACT 945

Search completed: September 3, 2000, 17:50:57

Job time: 13489 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 17:50:57 ; Search time 196.21 Seconds
(without alignments)
4158.179 Million cell updates/sec

Title: US-09-183-972-3
Perfect score: 3261
Sequence: 1 taaacgaagaggttatct.....tactatatgacataatcaat 3261

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	0.6	1399	1 Q89344	Human mglur7 clone
2	20	0.6	2265	1 Q24332	Mutant thermotabl
3	20	0.6	2403	1 Q24331	Mutant thermotabl
4	20	0.6	2568	1 Q28937	Encodes Taf DNA po
5	20	0.6	2571	1 Q24330	Mutant thermotabl
6	20	0.6	2679	1 Q24329	Mutant thermotabl
7	20	0.6	2679	1 Q28936	Encodes Asp37 Raf
8	20	0.6	3804	1 Q89343	Human mglur7 clone
9	20	0.6	4286	1 Q23917	Taf DNA polymerase
10	19	0.6	1301	1 X07172	Soybean diaminopim
11	18	0.6	1560	1 Q57071	AGE-modified DNA I
12	18	0.6	1893	1 Q05640	Clone hfs2 encodin
13	18	0.6	2070	1 Q82873	Human ST30 sialyt
14	18	0.6	2338	1 Q04774	Recombinant chole
15	18	0.6	2341	1 X13993	H. pylori GHP 675
16	18	0.6	2685	1 Q54029	Flocculation prote
17	18	0.6	2885	1 Q67260	S. cerevisiae FLO1
18	18	0.6	2754	1 T58505	R218 cDNA encoding
19	18	0.6	2754	1 V84039	HNA encoding rat H
20	18	0.6	2764	1 V52373	Streptococcus pneu
21	18	0.6	2844	1 N71250	Sequence of Vicia
22	18	0.6	4614	1 Q71390	Yeast 4.7 kb agglu
23	18	0.6	7812	1 X13962	Enterococcus faeca
24	18	0.6	35100	1 V20441	Human c-fms oncoge
25	18	0.6	110000	1 X20248_00	Borrelia burgdorfe
26	18	0.6	134525	1 Q04525	Total base sequenc
27	17	0.5	355	1 T19541	Human gene signatu
28	17	0.5	363	1 T90012	Heavy chain variab
29	17	0.5	523	1 X30918	Streptococcus pneu
30	17	0.5	550	1 T43492	ATM gene exon 60
31	17	0.5	641	1 Q94409	Aspergillus oryzae
32	17	0.5	666	1 T98554	DNA encoding a S.
33	17	0.5	730	1 X37499	Human secreted pro

Query Match 0.6%; Score 20; DB 1; Length 1399;

34 17 0.5 752 1 V53451 DNA encoding a Sta
35 17 0.5 920 1 Q32366 MAGE-10 genomic DN
36 17 0.5 920 1 Q72491 Tumour rejection a
37 17 0.5 920 1 T01170 MAGE-10 gene. Dete
38 17 0.5 1099 1 Q25831 Mutant human BCDP
39 17 0.5 1101 1 N90131 DNA sequence of pB
40 17 0.5 1101 1 N90135 DNA sequence of pB
41 17 0.5 1101 1 N90345 Plasmid pBSF2-L8,
42 17 0.5 1101 1 N80966 Plasmid pBSF2-L8,
43 17 0.5 1101 1 N81517 Sequence encoding
44 17 0.5 1128 1 Q74057 Human interleukin
45 17 0.5 1128 1 V60294 Human interleukin-

ALIGNMENTS

RESULT 1
Q89344 Q89344 standard; cDNA; 1399 BP.
AC Q89344;
DE 26-SEP-1995 (first entry)
DE Human mglur7 clone cmR3.
KW Human metabotropic glutamate receptor subtype 4; mglur7; hmglur7;
KW signal transducer; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
cds 1..270
FT misc_difference 920
FT /*tag= a
FT /*tag= b
FT /*note= "base n at position 920 is not identified
FT in the specification"
FT misc_difference 942
FT /*tag= c
FT /*note= "base n at position 942 is not identified
FT in the specification"
FT misc_difference 944
FT /*tag= d
FT /*note= "base n at position 944 is not identified
FT in the specification"
FT misc_difference 985
FT /*tag= e
FT /*note= "base n at position 985 is not identified
FT in the specification"
FT misc_difference 1090
FT /*tag= f
FT /*note= "base n at position 1090 is not identified
FT in the specification"

WO9508627-A.

PD 10-MAR-1995

PF 07-SEP-1994; E02991

PR 20-SEP-1993; EP-810663.

PR 19-AUG-1994; GB-016553.

PA (CIBA) CIBA GEIGY AG.

PI Flor PJ, Knoepfel T, Kuhn R, Lindauer K, Puettnner I;

DR WPI: 95-139596/18.

DR P-PSDB; R72094.

PT Human metabotropic glutamate receptor sub-type(s), esp. 4, 6 and
PT 7 - also corresp. DNA and antibodies, useful for identifying
PT cpds. which modulate signal transduction activity

PS Claim 17: Page 62-64; 110pp; English.

CC Human metabotropic glutamate receptor subtype 7 (hmglur7) cDNA

CC clones were isolated from cDNA libraries using a rat mglur4 probe.

CC Hippocampus partial cDNA clone cmR3, encoding the protein given in

CC R72094, was obtained. The missing 5' region of the clone was

CC generated by PCR from brain cDNA. Sequence comparison of fetal

CC brain and hippocampus clones indicated the existence of 2 subtypes

CC of hmglur7 (R72097-98).

SQ Sequence 1399 BP; 401 A; 280 C; 275 G; 438 T;

Query Match 0.6%; Score 20; DB 1; Length 2568;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3112 aataatcttgaattgtct 3131
 |||||
 Db 439 AATAATCTTTGAATTGTCT 420

RESULT 5

Q24330/c
 ID Q24330 standard; DNA; 2571 BP.
 AC Q24330;
 DT 22-OCT-1992 (first entry)
 DE Mutant thermostable DNA polymerase pTAFD2-37.
 KW 5'-3'; exonuclease; PCR; amplification; SSR; sequencing; PLCR; ss.
 OS Thermosipho africanus.
 FH Key Location/Qualifiers
 FT old_sequence 1..2
 FT /tag= a
 FT /note= "nucleotides 4-1111 deleted from the native
 sequence."
 FT
 PN WO9206200-A.
 PD 16-APR-1992.
 PF 30-SEP-1991; U07035.
 PR 28-SEP-1990; US-590213.
 PR 28-SEP-1990; US-590466.
 PR 28-SEP-1990; US-590490.
 PA (CETU) CETUS CORP.
 PI Abramson RD, Gelfand DH;
 DR WPI: 92-150885/18.
 DR P-PSDB; R23170.
 PT Thermostable DNA polymerases with altered 5'-3' exo nuclease
 PT activity - having conserved regions mutated or deleted, for use
 PT in e.g. PCR, sequencing and detection assays
 PS Claim 11; Page 59; 185pp; English.
 CC The sequence is that of Thermosipho africanus polymerase DNA which
 CC has been mutated. The mutation designated pTAFD2-37 causes the
 CC polymerase enzyme produced to exhibit a different amt. of 5'-3'
 CC exonuclease activity than the native enzyme. Thermostable DNA
 CC polymerases are useful in many recombinant DNA techniques, esp.
 CC nucleic acid amplification by PCR, self-sustained sequence
 CC replication (SSR) and high temp. DNA sequencing. The absence of
 CC 5'-3' nuclease activity may facilitate higher sensitivity allelic
 CC discrimination in a combined polymerase ligase chain reaction (PLCR)
 CC assay. An enhanced amt. of 5'-3' exonuclease activity may be desirable
 CC in enzymes used in homogeneous assays for the amplification and
 CC detection of a target nucleic acid sequence. Mutation of the DNA
 CC encoding particular regions of the enzymes can be used to prepare
 CC a range of recombinant proteins having 5'-3' exonuclease activity
 CC to a complete lack of activity.
 CC See also Q23993-Q24013, Q24320-36 and Q24343-60.
 SQ Sequence 2571 BP; 1014 A; 278 C; 494 G; 785 T;

Query Match 0.6%; Score 20; DB 1; Length 2571;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3112 aataatcttgaattgtct 3131
 |||||
 Db 442 AATAATCTTTGAATTGTCT 423

RESULT 6

Q24329/c
 ID Q24329 standard; DNA; 2679 BP.
 AC Q24329;
 DT 22-OCT-1992 (first entry)
 DE Mutant thermostable DNA polymerase from Thermosipho africanus.
 KW 5'-3'; exonuclease; PCR; amplification; SSR; sequencing; PLCR; ss.

OS Thermus thermophilus.
 FH Key Location/Qualifiers
 FT old_sequence 110
 FT /tag= a
 FT /note= "G in native sequence"
 FT old_sequence 111
 FT /tag= b
 FT /note= "A in native sequence"
 FT
 PN WO9206200-A.
 PD 16-APR-1992.
 PF 30-SEP-1991; U07035.
 PR 28-SEP-1990; US-590213.
 PR 28-SEP-1990; US-590466.
 PR 28-SEP-1990; US-590490.
 PA (CETU) CETUS CORP.
 PI Abramson RD, Gelfand DH;
 DR WPI: 92-150885/18.
 DR P-PSDB; R23169.
 PT Thermostable DNA polymerases with altered 5'-3' exo nuclease
 PT activity - having conserved regions mutated or deleted, for use
 PT in e.g. PCR, sequencing and detection assays
 PS Claim 11; Page 59; 185pp; English.
 CC The sequence is that of Thermosipho africanus polymerase DNA which has
 CC been mutated. The mutation causes the polymerase enzyme produced
 CC to exhibit a different amt. of 5'-3' exonuclease activity than the
 CC native enzyme. Thermostable DNA polymerases are useful in many
 CC recombinant DNA techniques, esp. nucleic acid amplification by
 CC PCR, self-sustained sequence replication (SSR) and high temp. DNA
 CC sequencing. The absence of 5'-3' nuclease activity may facilitate
 CC higher sensitivity allelic discrimination in a combined polymerase
 CC ligase chain reaction (PLCR) assay. An enhanced amt. of 5'-3'
 CC exonuclease activity may be desirable in enzymes used in homogeneous
 CC assays for the amplification and detection of a target nucleic acid
 CC sequence. Mutation of the DNA encoding particular regions of the
 CC enzymes can be used to prepare a range of recombinant proteins having
 CC 5'-3' exonuclease activity to a complete lack of activity.
 CC See also Q23993-Q24013, Q24320-36 and Q24343-60.
 SQ Sequence 2679 BP; 1045 A; 295 C; 515 G; 824 T;

Query Match 0.6%; Score 20; DB 1; Length 2679;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3112 aataatcttgaattgtct 3131
 |||||
 Db 550 AATAATCTTTGAATTGTCT 531

RESULT 7

Q28936/c
 ID Q28936 standard; DNA; 2679 BP.
 AC Q28936;
 DT 27-OCT-1992 (first entry)
 DE Encodes Asp37 Taf DNA polymerase I lacking 5'-3' exonuclease activity.
 KW Thermostability: PCR; polymerase chain reaction;
 KW thermophilic bacteria; Taf Pol I; mutant; ss.
 OS Thermosipho africanus.
 FH Key Location/Qualifiers
 FT mutation 109..111
 FT /tag= a
 FT /note= "Gly codon changed to Asp"

PN WO9206202-A.
 PD 16-APR-1992.
 PF 26-SEP-1991; U07076.
 PR 28-SEP-1990; US-590490.
 PA (CETU) CETUS CORP.
 PI Abramson RD, Gelfand DH, Greenfield L, Lawyer FC, Reichert FL;
 DR WPI: 92-150887/18.
 DR P-PSDB; R23122.
 PT Thermostable DNA polymerase from Thermosipho africanus - prepd.
 PT by purificn. from cells or by expression of Taf polymerase gene
 PT in host cells

PS Claim 20; Page 70; 80pp; English.
 CC The sequence coding for a thermostable DNA polymerase was isolated
 CC from chromosomal DNA of *Thermosipho africanus* (Taf). The polymerase
 CC (see R23122) has 5'-3' exonuclease activity. Mutation of the codon
 CC specifying Gly at position 37 (i.e. GGA) to an Asp codon results in
 CC a DNA polymerase which lacks the 5'-3' exonuclease activity.
 CC See Q23917 for the wild-type Taf Pol I gene and Q28937 for another
 CC preferred mutant.
 SQ Sequence 2679 BP; 1045 A; 295 C; 515 G; 823 T;

Query Match 0.6%; Score 20; DB 1; Length 2679;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3112 aataatctttgaattgtct 3131
 |||||
 DB 550 AATAATCTTTGAATTGTCT 531

RESULT 8
 Q89343
 ID Q89343 standard; cDNA; 3804 BP.
 AC Q89343;
 DT 26-SEP-1995 (first entry)
 DE Human mGluR7 clone cMR2.
 KW Human metabotropic glutamate receptor subtype 4; mGluR7; hmGluR7;
 KW signal transducer; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Key 1..2604
 FT cds /*tag= a
 FT misc_difference 3325 /*tag= b
 FT /note= "base n at position 3325 is not identified
 FT in the specification"
 FT misc_difference 3346 /*tag= c
 FT /note= "base n at position 3346 is not identified
 FT in the specification"
 FT misc_difference 3351 /*tag= d
 FT /note= "base n at position 3351 is not identified
 FT in the specification"
 FT misc_difference 3390 /*tag= e
 FT /note= "base n at position 3390 is not identified
 FT in the specification"
 FT misc_difference 3494 /*tag= f
 FT /note= "base n at position 3494 is not identified
 FT in the specification"

PN W09508627-A.
 PD 30-MAR-1995.
 PF 07-SEP-1994; E02991.
 PR 20-SEP-1993; EP-810663.
 PR 19-AUG-1994; GB-016553.
 PA (CIBA) CIBA GEIGY AG.
 PI Flor PJ, Knoepfel T, Kuhn R, Lindauer K, Puettnner I;
 DR WPI: 95-139596/18.
 DR P-PSDB: R72093.
 PT Human metabotropic glutamate receptor subtype 7 (hmGluR7) cDNA
 PT 7 - also corresp. DNA and antibodies, useful for identifying
 PT cpds. which modulate signal transduction activity
 PS Claim 17; Page 49-56; 110pp; English.
 CC Human metabotropic glutamate receptor subtype 7 (hmGluR7) cDNA
 CC clones were isolated from cDNA libraries using a rat mGluR4 probe.
 CC Fetal brain partial cDNA clone cMR2, encoding the protein given in
 CC R72093, was obtained. The missing 5' region of the clone was
 CC generated by PCR from brain cDNA. Sequence comparison of fetal
 CC brain and hippocampus clones indicated the existence of 2 subtypes
 CC of hmGluR7 (R72097-98).
 SQ Sequence 3804 BP; 1050 A; 892 C; 863 G; 994 T;

Query Match 0.6%; Score 20; DB 1; Length 3804;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1136 catggaaccatggagagga 1155
 |||||
 DB 2688 CATGGAACCATGGAGGAGGA 2707

RESULT 9
 Q23917/c
 ID Q23917 standard; DNA; 4286 BP.
 AC Q23917;
 DT 27-OCT-1992 (first entry)
 DE Taf DNA polymerase I coding sequence.
 KW Thermostability; PCR; polymerase chain reaction;
 KW thermophilic bacteria; Taf Pol I; ss.
 OS *Thermosipho africanus*.
 FH Key Location/Qualifiers
 FT Key 298..2976
 FT cds /*tag= a
 FT /product= Polymerase_I
 PN W09206202-A.
 PD 16-APR-1992.
 PF 26-SEP-1991; U07076.
 PR 28-SEP-1990; US-590490.
 PA (CETU) CETUS CORP.
 PI Abramson RD, Gelfand DH, Greenfield L, Lawyer FC, Reichert FL;
 DR WPI: 92-150887/18.
 DR P-PSDB: R23122.
 PT Thermostable DNA polymerase from *Thermosipho africanus* - prepd.
 PT by purificn. from cells or by expression of Taf polymerase gene
 PT in host cells
 PS Claim 8; Page 6; 80pp; English.
 CC Chromosomal DNA from *Thermosipho africanus* (Taf) was PCR-amplified
 CC with degenerate primers corresponding to the amino acid sequences
 CC of conserved regions of known thermostable polymerases. When
 CC specific PCR products of a similar size to the product generated
 CC using Taq chromosomal DNA were produced, the PCR fragments were
 CC cloned and sequenced. Fragments with sequences which encoded
 CC regions of amino acid homology to known thermostable polymerases
 CC were identified. The cloned PCR products were used as probes to
 CC screen a genomic Southern blot. The full-length Taf coding sequence
 CC was then compiled from various clones. See also Q23918-Q23961.
 SQ Sequence 4286 BP; 1623 A; 470 C; 847 G; 1346 T;

Query Match 0.6%; Score 20; DB 1; Length 4286;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3112 aataatctttgaattgtct 3131
 |||||
 DB 847 AATAATCTTTGAATTGTCT 828

RESULT 10
 X07172/c
 ID X07172 standard; cDNA; 1301 BP.
 AC X07172;
 DT 21-MAY-1999 (first entry)
 DE Soybean diaminopimelate epimerase cDNA.
 KW Diaminopimelate epimerase; soybean; amino acid; lysine;
 KW threonine; methionine; cysteine; isoleucine; transgenic plant;
 KW crop improvement; food; feedstuff; ss.
 OS Glycine max.
 FH Key Location/Qualifiers
 FT Key 84..1163
 FT cds /*tag= a
 PN W09855601-A2.
 PD 10-DEC-1998.

PF 05-JUN-1998; 011692.
 PR 12-JUN-1997; US-049443.
 PR 06-JUN-1997; US-048771.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PI Abell LM, Allen SM, Falco SC, Hitz WD, Kinney AJ,
 PI Rafalski JA, Thorpe CJ;
 DR P-PSDB; W97731.
 DR WPI; 99-070263/06.
 PR New plant amino acid biosynthetic enzymes, DNA and chimeric genes -
 PT encode: dihydropicolinate reductase; diaminopimelate epimerase;
 PT threonine synthase; threonine deaminase; S-adenosylmethionine
 PT synthetase
 PS Claim 7; Page 41-42; 98pp; English.
 CC This is the nucleotide sequence of a contig of cDNA clones
 CC se2.pk0005.fl, ses8w.pk0010.h11, sf11.pk0031.h3 and sgs1c.pk002.k12
 CC which codes for full-length soybean dihydropicolinate reductase
 CC (see W97731). The clones were isolated from soybean embryo (10
 CC days after flowering), mature embryo (8 weeks after subculture),
 CC immature flower and seed (4 hr after germination) cDNA libraries
 CC and identified by comparison to public sequence databases using
 CC BLAST algorithms. The contig shows sequence similarity to the
 CC Synchocystis sp. enzyme. The contig relates to new isolated
 CC nucleic acid fragments (see X07168-85) encoding plant enzymes (see
 CC W97727-44) that catalyze steps in the biosynthesis of lysine,
 CC threonine, methionine, cysteine and isoleucine from aspartate, the
 CC enzyme being selected from dihydropicolinate reductase,
 CC diaminopimelate epimerase, threonine synthase, threonine deaminase
 CC or S-adenosylmethionine synthetase. The invention also relates to
 CC the construction of a chimeric gene encoding all or a portion of
 CC the biosynthetic pathway enzyme, in sense or antisense orientation,
 CC where expression of the chimeric gene results in production of
 CC altered levels of the enzyme in a transformed host cell.
 CC Overexpression or reduction of expression of genes encoding the
 CC amino acid biosynthetic pathway enzymes in crop plants such as
 CC corn, soybean and wheat can be used to alter levels of the amino
 CC acids in human food and animal feed. Transformed host cells can
 CC also be used to identify compounds that inhibit one of the enzymes.
 SQ Sequence 1301 BP; 333 A; 291 C; 313 G; 364 T;

Query Match 0.6%; Score 19; DB 1; Length 1301;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 atttgcaagcatcgac 330
 Db 552 ATTGGCAAGCATCGAC 534

RESULT 11

ID Q57071 standard; DNA; 1560 BP.
 AC Q57071;
 DT 26-JUL-1994 (first entry)
 DE AGE-modified DNA INS-34.
 KW Advanced glycosylation end products; AGE plasmids; transposon; ss.
 OS Mus musculus.
 PN W03402599-A.
 PD 03-FEB-1994.
 PF 19-JUL-1993; U06754.
 PR 22-JUL-1992; US-920985.
 PA (UYRQ) UNIV ROCKEFELLER.
 PI Bucala RJ, Cerami A, Lee AT;
 DR WPI; 94-048857/06.
 PT Advanced glycosylation end-products, typically in the form of
 PT age-plasmids - can be transfected into cells and used to capture
 PT or activate transposons, e.g. to treat tumour cells
 PS Claim 6; Fig 3; 55pp; English.
 CC The DNA sequence comprises a portion of a transposon INS-20 that
 CC affects expression and related cellular activity. The DNA has been
 CC reacted with advanced glycosylation end products and is typically in
 CC the form of an AGE plasmid that can be transfected into cells. The
 CC AGE modification of the plasmid may activate the transposons which

CC are captured. Such capture or movement of transposons in a cell may
 CC be used to treat tumour cells.
 CC See also Q57059-73.
 SQ Sequence 1560 BP; 464 A; 316 C; 353 G; 427 T;
 Query Match 0.6%; Score 18; DB 1; Length 1560;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 858 accagaagtccaagcgag 875
 Db 1071 ACCAGAAGTTCAGGCAG 1088

RESULT 12

Q05640/c
 ID Q05640 standard; DNA; 1693 BP.
 AC Q05640;
 DT 17-DEC-1990 (first entry)
 DE Clone hps2 encoding SPL-like protein.
 KW Pregnancy-specific protein; carcinoembryonic antigen; CEA;
 KW fertility; abortion; placenta; ss.
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT cds 1..1065
 FT /*tag= a
 FT repeat_region 286..819
 FT /*tag= b
 FT repeat_unit 286..564
 FT /*tag= c
 FT /*label= Rn
 FT repeat_unit 566..819
 FT /*tag= d
 FT /*label= Rc
 FT poly_a_signal 1705..1710
 FT /*tag= e
 FT poly_a_site 1725..1753
 FT /*tag= f
 PN W09007937-A.
 PD 26-JUL-1990.
 PF 11-JAN-1990; U00285.
 PR 18-JAN-1989; US-298638.
 PR 07-AUG-1989; US-390409.
 PA (OKLA-) OKLAHOMA MED RES FO.
 PI Chan WY;
 PI WPI; 90-253860/33.
 DR P-PSDB; R96429.
 PT Pregnancy specific proteins, genes and antibodies - for use in
 PT diagnosis and in compans. having immunosuppressive and growth
 PT promoting activities.
 PS Claim 4; Fig 3; 78pp; English.
 CC hps2 is a clone encoding a placental pregnancy-specific protein
 CC (SPL) also known as pregnancy-specific beta glycoprotein (PSBG)
 CC which is detected in placenta and is membrane-bound. The sequence
 CC and Abs specific for it can be used in diagnosis, pregnancy
 CC testing and monitoring of tumours. The protein has immuno-
 CC suppressive activity and growth promoting activity and can be used
 CC for enhancing fertility in females, viability of a foetus etc.
 CC See also Q05639-46.
 SQ Sequence 1693 BP; 440 A; 432 C; 398 G; 423 T;

Query Match 0.6%; Score 18; DB 1; Length 1693;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 atataataattattttcc 54
 Db 129 ATATATAATTATTTTCC 112

RESULT 13


```

Q82873
ID Q82873 standard; cDNA; 2070 BP.
AC Q82873;
DE 27-SEP-1995 (first entry)
DT Human ST30 sialyltransferase.
KW Sialyltransferase; sialic acid; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 931..1953
FT /*tag= a
FN WO9504816-A.
PD 16-FEB-1995.
PE 27-JUL-1994; U08516.
PR 04-AUG-1993; US-102385.
PA (REGC ) UNIV CALIFORNIA.
PI Burlingame AL, Gillespie W, Kelm S, Livingston B;
PI Medzhiradzky K, Paulson JC, Wen X;
DR WPI; 95-090894/12.
DR P-PSDB; R65244.
PT Prod. of mammalian sialyltransferase(s) - useful in the
PT addition of sialic acids on carbohydrate(s) and the
PT identification of other members of the same gene family
PS Claim 66; Figure 20; 136pp; English.
CC Human Gal beta 1,3GalNAc alpha 2,3 sialyltransferase is also
CC called human ST30. In order to clone human ST30 sialyltransferase
CC two degenerate oligos were synthesised (Q82880 & Q82881). For
CC PCR amplification, first strand cDNA synthesised from human
CC placenta or human fetal brain total RNA was combined with each
CC primer. 8/50 clones obtd. from human placenta were judged to
CC contain the human ST30 sialylmotif as judged by homology with the
CC porcine sequence. A human placenta cDNA library was screened with
CC the cloned PCR fragment. Characterisation of the positive clones
CC revealed cDNA of two types which differed in their 5' ends. The nt
CC sequence of the short form had a deletion from nt -253 - -37 and is
CC shown in Q82873.
SQ Sequence 2070 BP; 496 A; 567 C; 562 G; 445 T;

Query Match 0.6%; Score 18; DB 1; Length 2070;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3136 aagaaactgaggttcaga 3153
|||||
DB 303 AAGAACTGAGGTTTCAGA 320

RESULT 14
Q04774
ID Q04774 standard; DNA; 2338 BP.
AC Q04774;
DE 19-OCT-1990 (first entry)
DT Recombinant cholesterol oxidase gene.
KW Cholesterol oxidase; pSL81; cholesterol; ds.
OS Rhodococcus sp.
FH Key Location/Qualifiers
FT cds 146..2026
FT /*tag= a
FN WO9005788-A.
PD 31-MAY-1990.
PE 14-NOV-1989; 005112.
PR 14-NOV-1988; US-269669.
PA (GEN2-) Genzyme Corp.
PI Long S, Ostroff GR;
DR WPI; 90-193438/25.
DR P-PSDB; R05534.
PT Cloning vector contg. DNA from nocardioform microorganisms -
PT deoxyribonucleic acid.
PS Disclosure.
CC Cholesterol oxidase isolated from a nocardioform organism is more
CC stable, more active and free of lipids, it is expressed as an
CC extracellular protein and does not require an inducer.
CC The gene product can be used in assaying cholesterol by determining

```

```

CC the extent of oxidation.
SQ Sequence 2338 BP; 433 A; 878 C; 640 G; 387 T;

Query Match 0.6%; Score 18; DB 1; Length 2338;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1881 ccttctccaagacctgt 1898
|||||
DB 1275 CCTTCTCCAAGACCTGT 1292

RESULT 15
X13993/C
ID X13993 standard; DNA; 2341 BP.
AC X13993;
DT 31-MAR-1999 (first entry)
DE H. pylori GHPO 675 gene.
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT cds 966..2294
FT /*tag= a
FN W09843478-AL.
PD 08-OCT-1998.
PE 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI; 98-542293/46.
DR P-PSDB; W98274.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 297-299; 2054pp; English.
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 2341 BP; 1005 A; 238 C; 459 G; 639 T;

Query Match 0.6%; Score 18; DB 1; Length 2341;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 tttttaatgagatttgag 82
|||||
DB 93 TTTTAAATGAGATTTCAG 76

Search completed: September 3, 2000, 17:51:42
Job time: 13534 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 17:47:25 ; Search time 124.08 seconds
(without alignments)
3615.164 Million cell updates/sec

Title: US-09-183-972-3
Perfect score: 3261
Sequence: 1 taacacgaaggtatct.....tactatgacataatcaat 3261

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 243080 seqs, 68777915 residues

Word size : 0
Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	0.6	2679	1	US-07-977-434-11
C 2	20	0.6	2679	2	US-08-458-819-11
C 3	20	0.6	2679	6	PCT-US91-07035-11
C 4	18	0.6	1893	7	5169835-3
C 5	18	0.6	2885	5	US-08-362-525-21
C 6	18	0.6	2754	1	US-08-196-989B-1
C 7	18	0.6	2754	3	US-08-760-936-1
C 8	18	0.6	4614	1	US-08-325-267A-1
C 9	18	0.6	35100	2	US-08-306-691B-19
C 10	18	0.6	35100	6	PCT-US93-06251-19
C 11	17	0.5	920	1	US-08-299-849B-22
C 12	17	0.5	920	4	US-08-142-368A-22
C 13	17	0.5	920	5	US-08-967-727-22
C 14	17	0.5	1145	7	5510472-1
C 15	17	0.5	1534	2	US-08-592-126-97
C 16	17	0.5	1881	5	US-09-235-246-2
C 17	17	0.5	2352	3	US-08-889-909A-21
C 18	17	0.5	2559	3	US-08-724-774B-3
C 19	17	0.5	2651	3	US-08-786-164-5
C 20	17	0.5	2746	3	US-08-576-165-3
C 21	17	0.5	2757	3	US-08-627-254C-26
C 22	17	0.5	6545	6	PCT-US95-13749-3
C 23	17	0.5	10409	5	US-08-772-440-33
C 24	17	0.5	46899	2	US-08-471-119A-1
C 25	16	0.5	37	6	PCT-US91-02942-78
C 26	16	0.5	38	4	US-08-857-946-137

27	16	0.5	38	5	US-08-970-740-137	Sequence 137, Appl
c 28	16	0.5	76	2	US-08-657-012-6	Sequence 6, Appl
29	16	0.5	116	1	US-07-634-278-96	Sequence 96, Appl
30	16	0.5	116	1	US-08-477-728-96	Sequence 96, Appl
31	16	0.5	116	2	US-08-474-040-96	Sequence 96, Appl
32	16	0.5	116	2	US-08-487-200-96	Sequence 96, Appl
33	16	0.5	130	2	US-08-355-888A-13	Sequence 13, Appl
34	16	0.5	130	3	US-08-693-697-13	Sequence 13, Appl
35	16	0.5	130	4	US-08-640-389A-6	Sequence 6, Appl
c 36	16	0.5	288	1	US-08-410-804-22	Sequence 22, Appl
c 37	16	0.5	288	2	US-08-259-514-22	Sequence 22, Appl
c 38	16	0.5	288	3	US-08-858-311-22	Sequence 22, Appl
39	16	0.5	305	4	US-07-938-154-7	Sequence 7, Appl
40	16	0.5	305	6	PCT-US91-02311-7	Sequence 7, Appl
41	16	0.5	344	4	US-08-888-366-7	Sequence 7, Appl
42	16	0.5	348	3	US-08-273-146-54	Sequence 54, Appl
43	16	0.5	350	4	US-07-938-154-8	Sequence 8, Appl
44	16	0.5	350	6	PCT-US91-02311-8	Sequence 8, Appl
45	16	0.5	360	3	US-08-002-324-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-07-977-434-11/c
; Sequence 11, Application US/07977434
; Patent No. 5466591
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: Wordperfect 2.1
; CURRENT APPLICATION DATA: US/07977,434
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US90/07641
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cserr
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2679 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Thermosipho africanus
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2676
US-07-977-434-11

Query Match 0.68; Score 20; DB 1; Length 2679;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3112 aataatctttgaattgttct 3131
|||||
DB 550 AATAATCTTTGAATTGTCT 531

RESULT 2
US-08-458-819-11/c
Sequence 11, Application US/08458819
Patent No. 5795762
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Abramson, Richard D.
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 7
SOFTWARE: Wordperfect 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,819
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/977,434
FILING DATE: 23-FEB-1993
APPLICATION NUMBER: US 590,490
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,466
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,213
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,394
FILING DATE: 15-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 143,441
FILING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 063,509
FILING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FILING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US90/07641
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cserr
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2679 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Thermosipho africanus
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2676
US-08-458-819-11

Query Match 0.68; Score 20; DB 2; Length 2679;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3112 aataatctttgaattgttct 3131
|||||
DB 550 AATAATCTTTGAATTGTCT 531

RESULT 3

PCT-US91-07035-11/c
; Sequence 11, Application PC/TUS9107035
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abranson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cetus Corporation
; STREET: 1400 Fifty-third Street
; City: Emeryville
; STATE: California
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/07035
; FILING DATE: 19910930
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Sias Ph.D, Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: Case No. 2580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-420-3300
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2679 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomlc)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Thermosipho africanus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2676
; PCT-US91-07035-11

Query Match 0.6%; Score 20; DB 6; Length 2679;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3112 aataatctttgaattgttct 3131
|||||
Db 550 AATAATCTTTGAATTGTCT 531
|||||

RESULT 4
5169835-3/c
; Patent No. 5169835
; APPLICANT: WAI-YEE, CHAN
; TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS APPLICATIONS
; NUMBER OF SEQUENCES: 48
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/390,409
; FILING DATE: 07-AUG-1989
; SEQ ID NO:3
; LENGTH: 1693
5169835-3

Query Match 0.6%; Score 18; DB 7; Length 1693;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 atataataattttttcc 54
|||||
Db 129 ATATATAATTATTTTCC 112
|||||

RESULT 5
US-08-362-525-21
; Sequence 21, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2685 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: pY105
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2685
; OTHER INFORMATION: /product= "Flocculation protein" /gene= nFLO1"
; US-08-362-525-21

Query Match 0.6%; Score 18; DB 5; Length 2685;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ctgtgattgttcagaa 106
|||||
DB 1058 CTGTGATTGTTCAGAA 1075

RESULT 6
US-08-196-989B-1/c
; Sequence 1, Application US/08196989B
; Patent No. 5585476
; GENERAL INFORMATION:
; APPLICANT: MacLennan, A. John
; TITLE OF INVENTION: Molecular Cloning and Expression of
; TITLE OF INVENTION: G-Protein Coupled Receptors
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/196,989B
; FILING DATE: 15-FEB-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: MAC-100
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-196-989B-1

Query Match 0.6%; Score 18; DB 1; Length 2754;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 ttacgaattccaggag 574
|||||
DB 2560 TTCAGCAATTCAGGAG 2543

RESULT 7
US-08-760-936-1/c
; Sequence 1, Application US/08760936
; Patent No. 5856443
; GENERAL INFORMATION:
; APPLICANT: MacLennan, A. John
; TITLE OF INVENTION: Molecular Cloning and Expression of
; TITLE OF INVENTION: G-Protein Coupled Receptors
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,936
; FILING DATE: December 6, 1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: MAC-100C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-760-936-1

Query Match 0.6%; Score 18; DB 3; Length 2754;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 ttacgaattccaggag 574
|||||
DB 2560 TTCAGCAATTCAGGAG 2543
```

```
RESULT 8
US-08-325-267A-1
; Sequence 1, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TARATA, YOSHIIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENNITILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-1D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4614
US-08-325-267A-1

Query Match 0.6%; Score 18; DB 1; Length 4614;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ctgtgattgtatcagaa 106
|||||
DB 2948 CTGTGATTGTATCAGAA 2965

RESULT 9
US-08-306-691B-19
; Sequence 19, Application US/08306691B
; Patent No. 5734039
```

```
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
; STREET: Two Penn Center, suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-306-691B-19

Query Match 0.6%; Score 18; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2113 tgcgtcagcccaacaact 2130
|||||
DB 34227 TGCTGAGCCCAACAAC 34244

RESULT 10
PCT-US93-06251-19
; Sequence 19, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
```

ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-19

Query Match 0.6%; Score 18; DB 6; Length 35100;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2113 tgctgcagcccaact 2130
|||||
Db 34227 TGCTGAGCCCAACT 34244

RESULT 11
US-08-299-849B-22/c
Sequence 22, Application US/08299849B
Patent No. 5612201
GENERAL INFORMATION:
APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
APPLICANT: Chomez, Patrick
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
Determining Expression Of A Tumor Antigen Precursor
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,849B
FILING DATE: 1-SEPTEMBER-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,230
FILING DATE: 26-MARCH-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991

ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5612201man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5355
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-10 gene
US-08-299-849B-22

Query Match 0.5%; Score 17; DB 1; Length 920;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2277 acagccaggggagcctg 2293
|||||
Db 434 ACAGCCAGGGGAGCCTG 418

RESULT 12
US-08-142-368A-22/c
Sequence 22, Application US/08142368A
Patent No. 5925729
GENERAL INFORMATION:
APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
Rejection Antigen Precursors, Tumor
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/142,368A
FILING DATE: 02-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5925729man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5253.4-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-10 gene
US-08-142-368A-22

Query Match 0.5%; Score 17; DB 4; Length 920;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2277 acagccagggagcctg 2293
|||||
Db 434 ACAGCCAGGGGAGCCTG 418

RESULT 13
US-08-967-727-22/c
Sequence 22, Application US/08967727
Patent No. 6025474
GENERAL INFORMATION:
APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
APPLICANT: van der Brugten, Pierre; Boon-Falleur, Thierry
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felre & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,727
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,230
FILING DATE: 26-MARCH-1993
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,365
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6025474man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-10 gene
US-08-967-727-22

Query Match 0.5%; Score 17; DB 5; Length 920;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2277 acagccagggagcctg 2293
|||||
Db 434 ACAGCCAGGGGAGCCTG 418

RESULT 14
5510472-1
Patent No. 5510472
APPLICANT: REVEL, MICHEL; TIOLLAIS, PIERRE
TITLE OF INVENTION: PRODUCTION OF RECOMBINANT HUMAN
INTERFERON-BETA2
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/883,633
FILING DATE: 15-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 208,925
FILING DATE: 20-NOV-1980
APPLICATION NUMBER: 449,447
FILING DATE: 12-DEC-1989
APPLICATION NUMBER: 860,883
FILING DATE: 08-MAY-1986
APPLICATION NUMBER: 208,925
FILING DATE: 20-NOV-1980
SEQ ID NO: 1
LENGTH: 1145
5510472-1

Query Match 0.5%; Score 17; DB 7; Length 1145;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1754 ccagatcattcttggg 1770
|||||
Db 1014 ccagatcattcttggg 1030

RESULT 15
US-08-592-126-97
Sequence 97, Application US/08592126
Patent No. 5821091
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
TITLE OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

Search completed: September 3, 2000, 17:49:21
Job time: 17200 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 10:24:07 ; Search time 2738.16 Seconds
(without alignments)
5251.866 Million cell updates/sec

Title: US-09-183-972-3
Perfect score: 3261
Sequence: 1 taacaagaaggtatctct.....tactatgacataatcaat 3261

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

```
117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
C	1	459.6	14.1	561	91	W26960	W26960 16h10 Human
	2	354.4	10.9	405	85	H38604	H38604 yp48e04_r1
	3	317	9.7	322	24	AA326863	AA326863 EST30113
	4	298.8	9.2	309	85	H38594	H38594 yp48c06.r1
	5	291	8.9	298	24	AA96278	AA96278 EST10795
	6	200.8	6.2	481	40	AI510373	AI510373 mp95e12.y
	7	200.8	6.2	495	21	AA116463	AA116463 mp95e12.r
	8	90.6	2.8	600	122	FR0033795	AL030162 Fugu rubr
	9	81	2.5	647	95	AQ234080	AQ234080 HS_2057_A
	10	79	2.4	830	96	AQ253330	AQ253330 HS_2046_B
C	11	60.6	1.9	996	122	CNS00FUH	AL071063 Drosophill
	12	59.6	1.8	843	122	CNS00CS1	AL059666 Drosophill
	13	59.2	1.8	410	122	FR0033806	AL030173 Fugu rubr
	14	59.2	1.8	1101	122	CNS00HX9	AL073856 Drosophill
	15	59.2	1.8	1201	123	CNS0161H	AL106175 Drosophill
	16	58.4	1.8	1101	122	CNS00210	AL061917 Drosophill
	17	58.2	1.8	791	122	CNS009KS	AL053801 Drosophill
	18	57	1.7	867	122	CNS00CX5	AL060052 Drosophill
	19	57	1.7	1101	122	CNS00Z62	AL097301 Drosophill
	20	56.8	1.7	3707	81	C83838	C83838 C83838 Diet
C	21	56.4	1.7	1038	123	CNS0146B	AL103757 Drosophill
	22	55.8	1.7	1101	122	CNS00LT2	AL078714 Drosophill
	23	55.2	1.7	1101	122	CNS00LFB	AL060732 Drosophill
	24	55	1.7	614	123	CNS0152H	AL104915 Drosophill
	25	54.8	1.7	822	114	AQ752069	AQ752069 HS_5570_B
	26	54.6	1.7	550	103	AQ0500506	AQ050050 V382 mtN
	27	54.2	1.7	927	123	CNS016EF	AL106641 Drosophill
	28	54	1.7	884	122	CNS0129A	AL101272 Drosophill
	29	53.8	1.6	1101	122	CNS00EVL	AL069706 Drosophill
	30	53.4	1.6	1101	122	CNS0039G	AL063921 Drosophill
C	31	53.2	1.6	684	120	B10032	B10022 T31K5-Sp6.1
	32	53.2	1.6	832	101	AQ325582	AQ325582 nbx00015F
	33	53.2	1.6	1068	122	CNS00ETV	AL069846 Drosophill
	34	53.2	1.6	1141	123	CNS014AQ	AL103916 Drosophill
	35	53	1.6	447	43	AI739559	AI739559 w435a06.x
	36	53	1.6	833	122	CNS007X3	AL050945 Drosophill
	37	53	1.6	1043	123	CNS0145P	AL103735 Drosophill
	38	53	1.6	1101	122	CNS012NV	AL101797 Drosophill
	39	52.8	1.6	593	122	CNS00880	AL051540 Drosophill
	40	52.6	1.6	1101	122	CNS00LJT	AL068307 Drosophill
C	41	52.6	1.6	1101	122	CNS01219	AL101595 Drosophill
	42	52.4	1.6	783	122	CNS00AIS	AL055833 Drosophill
	43	52.4	1.6	1128	120	B11263	B11263 F35L19-T7 I
	44	52.2	1.6	467	122	CNS0138G	AL109298 Drosophill
	45	52.2	1.6	661	122	CNS013JT	AL102947 Drosophill

ALIGNMENTS

RESULT	1.
W26960/c	
LOCUS	W26960 561 bp mRNA EST
DEFINITION	16h10 Human retina cDNA randomly primed sublibrary Homo sapiens cDNA, mRNA sequence.


```

||||| 11 |||| ||||||| ||||||| ||||||| || |||| ||||| |||||
Db 122 TAAAGATAGTGGACAGAGGAAGCAGAGTGTCTGTCAGACAGGAGCATGAGACCCAG 181

QY 2286 ggagccggagggctgtggaaccagccctctgtggcctgagcacaaggaatgcaggtcct 2345
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 182 GGACCTGGACTACAGACCTGAACCTCTGTCCCTCTG--GAAAGACTTGTGTGCCGG 239

QY 2346 ccagggaaaggaggtccatgc--ggttccagatcactctgaaatcaagatacaaaaac 2403
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 240 CCGAGAACAGCAACTCCATCGAGGCCACCGAGTACATCTACAAACCAAGCTCAGGAAC 299

QY 2404 tagtgtaaaagttccaaaatacaaaaatacaaaaggttaatacagtaaaagaattctgaa 2463
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 300 TGGTGTAAAAA-----ACGTACGTGAGCAAAATAAGGTAGTCAAGAAAAGAAATCTTAA 354

QY 2464 ttactgcgttagaataatgaagaatttaaccatcaagattgggaaggaattaaaaactg 2523
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 355 CTATCAGCTATAGGATTTGAAGAATTTGAAGACCAGGACTGGGAGGAAATTAAGAAGCTG 414

QY 2524 aaaatgta 2531
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 415 GAATCATA 422

RESULT 7
AAL16463
LOCUS mp95e12.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:577006
DEFINITION 5', mRNA sequence.
EST
ACCESSION AAL16463
VERSION AAL16463.1 GI:1671475
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 495)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:351654
Seq primer: -28M13 rev2 from Amer sham
High quality sequence stop: 491.
Location/Qualifiers
1. .495
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:577006"
/clone_lib="Soares_thymus_2NbMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(df) primer [5'
TGTACCAATCTGAAGTGGGAGCGCGCGCTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not I

```

```

and Eco RI sites of the modified pMT3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 156 a 114 c 118 g 107 t
ORIGIN

Query Match 6.2%; Score 200.8; DB 21; Length 495;
Best Local Similarity 71.7%; Pred. No. 2e-41;
Matches 307; Conservative 0; Mismatches 112; Indels 9; Gaps 3;

QY 2106 ttctgttctgtcagcccaacacccatctctggaatagacagctactctctcaacattg 2165
Db 2 TTCGGTCCACCGCAGCTCAAGGCTCAATCTGGAATCGAAAAGCTACTCCCTCGACATTG 61

QY 2166 aaccagctatcaagcagatccctgcaagtcttctgctgcgcgcgaatttgcccaattg 2225
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 62 AACCCAGCTGATCAGCGGGATCCCTGCAAACTCTAGACTGTGCAAAATTTGCCAGTGTG 121

QY 2226 taagaacgaacgagctgaggaagcgagtgctcgtcgcaaacacagcagatgacacagcag 2285
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 122 TAAAGAAATCAGTGGACAGAGAGAGCAGAGTGTCTGCTGACAGAGGACATGAGAGCCAG 181

QY 2286 ggagcctggagcgtgtggaaccagcgctctgtgctggtgcacaaaaggaatgcaggtcct 2345
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 182 GGACCTGGACTACAGACCCCTGAACCTCTGTCCCTCTG--GAAAGACTTGTGTGCCGG 239

QY 2346 ccagggaaaggaggtccatgc--ggttccagatcactctgaaatcaagcagcagcagcag 2403
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 240 CCGAGAACAGCAACTCCATGTCAGGCCACAGATCTACTACAAACCAAGCTCAGGAAC 299

QY 2404 tagtgtaaaagtccaaaatacaaaaatacaaaaggttaatacagtaaaagaattctgaa 2463
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 300 TGGTGTAAAAA-----AGCTACGTGACCAAAATAAGGTAGTCAAGAAAAGAAATTTTAA 354

QY 2464 ttactgacctagaataatgaagaatttaaccatcaagattgggaaggaattaaaaactg 2523
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 355 CTATCAGCTATAGGATTTGAAGAATTTGAAGACCAGGACTGGGAGGAAATTAAGAAGCTG 414

QY 2524 aaaatgta 2531
||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
Db 415 GAATCATA 422

RESULT 8
FR0033795
LOCUS Fugu rubripes GSS sequence, clone 184G17AD12, genomic survey
DEFINITION sequence.
ACCESSION AL030162
VERSION AL030162.1 GI:3272276
KEYWORDS GSS; genome survey sequence.
SOURCE Fugu rubripes.
ORGANISM Fugu rubripes
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrana,Y.,
Williams,G. and Brenner,S.
Direct Submission
Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:
biohelp@hgmrc.ac.uk
Vector: pBluescript II KS
V.type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
FEATURES
Location/Qualifiers
1. .600
source

```



```
/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 184g17"
/clone="184G17ad12"
BASE COUNT      126 a   141 c   181 g   108 t   44 others
ORIGIN

Query Match      2.8%; Score 90.6; DB 122; Length 600;
Best Local Similarity 67.6%; Pred. No. 9.1e-13;
Matches 123; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1948 gctgctgttcatactcagatcccaatcttacaggatttaagcaacttgaaataacttaa 2007
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 GCAGCTGCTGCATACCTGCGNGCCCACTGACGGGATNCAAGAACCTGGAGATCCTGAA 88

QY 2008 ctctcagaacaggagtggtgatttgtaagtaagcaaaatgaagtgtgctgacgcta 2067
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 CTTACGAGGAGGAGTGCTGCTCAACAGCAAGGTCAAAATTTGCCAAGATGGTGCNTA 148

QY 2068 taacctcaccaggctgtgcacaggggtcttgaggatttcgtctctcagcccaaca 2127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 149 CAACGTCACGGAGGCCGCTCGCTGCTGGANGAGTCTGCTCCACCGCCTCCAANGA 208

QY 2128 ac 2129
      ||
DB 209 AC 210

RESULT 9
LOCUS      AQ234080      647 bp      DNA      GSS      29-SEP-1998
DEFINITION HS_2057_AL_A05_MR CIT Approved Human Genomic Sperm Library D
            Homo sapiens genomic clone Plate=2057 Col=9 Row=A, genomic
            survey sequence.
ACCESSION  AQ234080

VERSION    AQ234080.1 GI:3662687
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 647)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2057 row: A column: 9
            Class: BAC ends
            High quality sequence stop: 647.
            Location/Qualifiers
                source
                    1..647
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Plate=2057 Col=9 Row=A"
                        /clone_lib="CIT Approved Human Genomic Sperm Library D"
                        /sex="male"
                        /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
                        E-Coli DH10B"
BASE COUNT      199 a   149 c   106 g   182 t   11 others
ORIGIN

/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 184g17"
/clone="184G17ad12"
BASE COUNT      126 a   141 c   181 g   108 t   44 others
ORIGIN

Query Match      2.8%; Score 90.6; DB 122; Length 600;
Best Local Similarity 67.6%; Pred. No. 9.1e-13;
Matches 123; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1948 gctgctgttcatactcagatcccaatcttacaggatttaagcaacttgaaataacttaa 2007
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 GCAGCTGCTGCATACCTGCGNGCCCACTGACGGGATNCAAGAACCTGGAGATCCTGAA 88

QY 2008 ctctcagaacaggagtggtgatttgtaagtaagcaaaatgaagtgtgctgacgcta 2067
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 CTTACGAGGAGGAGTGCTGCTCAACAGCAAGGTCAAAATTTGCCAAGATGGTGCNTA 148

QY 2068 taacctcaccaggctgtgcacaggggtcttgaggatttcgtctctcagcccaaca 2127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 149 CAACGTCACGGAGGCCGCTCGCTGCTGGANGAGTCTGCTCCACCGCCTCCAANGA 208

QY 2128 ac 2129
      ||
DB 209 AC 210

RESULT 9
LOCUS      AQ234080      647 bp      DNA      GSS      29-SEP-1998
DEFINITION HS_2057_AL_A05_MR CIT Approved Human Genomic Sperm Library D
            Homo sapiens genomic clone Plate=2057 Col=9 Row=A, genomic
            survey sequence.
ACCESSION  AQ234080

VERSION    AQ234080.1 GI:3662687
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 647)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2057 row: A column: 9
            Class: BAC ends
            High quality sequence stop: 647.
            Location/Qualifiers
                source
                    1..647
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Plate=2057 Col=9 Row=A"
                        /clone_lib="CIT Approved Human Genomic Sperm Library D"
                        /sex="male"
                        /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
                        E-Coli DH10B"
BASE COUNT      199 a   149 c   106 g   182 t   11 others
ORIGIN
```

```
Query Match      2.5%; Score 81; DB 95; Length 647;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1338 aggatgctactttagtccagaacttcctctgttgtaaccccaagcttgagacagtggacg 1397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 355 AGGATGCTACTTTGAGTCCAGAACTTCCTCTGTTGTAACCCAGCTTGAGACAGTGGACG 414

QY 1398 gagcagagcatggtctacctg 1418
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 415 GAGCAGAGCATGGTCTACCTG 435

RESULT 10
LOCUS      AQ253330      830 bp      DNA      GSS      10-OCT-1998
DEFINITION HS_2046_B1_H03_MR CIT Approved Human Genomic Sperm Library D
            Homo sapiens genomic clone Plate=2046 Col=5 Row=P, genomic
            survey sequence.
ACCESSION  AQ253330

VERSION    AQ253330.1 GI:3724684
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 830)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2046 row: P column: 5
            Class: BAC ends
            High quality sequence stop: 830.
            Location/Qualifiers
                source
                    1..830
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Plate=2046 Col=5 Row=P"
                        /clone_lib="CIT Approved Human Genomic Sperm Library D"
                        /sex="male"
                        /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
                        E-Coli DH10B"
BASE COUNT      122 a   436 c   80 g   191 t   1 others
ORIGIN

Query Match      2.4%; Score 79; DB 96; Length 830;
Best Local Similarity 89.5%; Pred. No. 1e-09;
Matches 85; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1014 gctcaagctccacagagatgcaacttaaggccatctttaagacacacagtgcagagcaa 1073
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 120 GCTCAAGCTTCACAGATGCAACTTACCGCCATCTTTAAGACACACAGCGCAGAACCA 179

QY 1074 aaagccctgcagtgacctctctgtcttttgattcc 1108
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 180 AAAGCCCGCAAGTGACCCACTGCTTTTGCTCC 214
```

```

RESULT 11
CNS000FUH      996 bp      DNA      GSS      03-JUN-1999
LOCUS          Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION     BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit
                fly), genomic survey sequence.
ACCESSION      AL071063
VERSION        AL071063.1 GI:4951105
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 996)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
COMMENT        Determination of this BAC-end sequence was carried out as part of a
                collaboration with the Berkeley Drosophila Genome Project (BDGP).
                The BDGP is constructing a physical map of the Drosophila
                melanogaster genome using these BACs. For further information
                please see http://www.fruitfly.org The BDGP Drosophila
                melanogaster BAC library was prepared by Kazutoyo Osoegawa and
                Aaron Mammoser in Pieter de Jong's laboratory in the Department of
                Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                NY. The library is named RPCI-98 and was constructed by partial
                EcoRI digestion of Drosophila DNA provided by the BDGP from the
                isogenic strain y2; cn bw sp, the same strain used for the BDGP's
                P1 and EST libraries. A more detailed description of the library
                and how to order individual BAC clones, the entire library, or
                filters for hybridization from the BACPAC Resource Center can be
                found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
    source
        1..996
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="RPCI-98"
            /clone="BACR31021"
            /note="end : TET3"

BASE COUNT      383 a 164 c 81 g 171 t 197 others
ORIGIN

Query Match      1.9%; Score 60.6; DB 122; Length 996;
Best Local Similarity 34.4%; Pred. No. 6.7e-05;
Matches 188; Conservative 83; Mismatches 275; Indels 1; Gaps 1;

QY 2355 gggagctccatggttcagatcactctgaaatcaagcagcacaactagtgttaaa 2414
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 KBBBCCCCCCCCCYCMHNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 510

QY 2415 gtcccaaaatacaaaatacaaggttaatacagtaaaagaattctgaattactgac 2474
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 ACGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 570

QY 2475 agaataagaataattacacatcaagattgggaaggaataaaactgaaatgtacaa 2534
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 WAAAWTAAATATATTTTAAAAAAAWWTTTCAAAATTTAAATTTAAATTTAAAT 630

QY 2535 ttatcattaggctatctcaagagagatgattgctctcaaggaaatggagacagc 2594
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 631 TWATTATTAAYWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 690

QY 2595 atattcatgggtctcaaaatcccgacatacagtcacactgagaatcagcacacac 2654
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 AAATWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 749

QY 2655 atttcaaatatagagatgctgactgctgcaaccagtaattctgaaataaagac 2714
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 750 AATTWATMAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 809
QY 2715 tactattataaaaccccaatcgagcgaacatatttttactattcttgatga 2774
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 810 TWAAATWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 869
QY 2775 tagtcaaatgatcataagccaggtttgtctccaccttccctgaaatatttactcacaga 2834
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 870 YTTAAATWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 929
QY 2835 tcatttgcacaagcatagcttactattgttaggactgacacattatttgggaagca 2894
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 930 AYAATTAATWAAATKTWATTTWATTTTAAATWAAAAAATAAATAAATAAATA 989
QY 2895 aactctt 2901
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 990 AAWTTTT 996

RESULT 12
CNS000CS1/c    843 bp      DNA      GSS      04-JUN-1999
LOCUS          Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION     BACR26H19 of RPCI-98 library from Drosophila melanogaster (fruit
                fly), genomic survey sequence.
ACCESSION      AL059666
VERSION        AL059666.1 GI:4947129
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 843)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
COMMENT        Determination of this BAC-end sequence was carried out as part of a
                collaboration with the Berkeley Drosophila Genome Project (BDGP).
                The BDGP is constructing a physical map of the Drosophila
                melanogaster genome using these BACs. For further information
                please see http://www.fruitfly.org The BDGP Drosophila
                melanogaster BAC library was prepared by Kazutoyo Osoegawa and
                Aaron Mammoser in Pieter de Jong's laboratory in the Department of
                Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                NY. The library is named RPCI-98 and was constructed by partial
                EcoRI digestion of Drosophila DNA provided by the BDGP from the
                isogenic strain y2; cn bw sp, the same strain used for the BDGP's
                P1 and EST libraries. A more detailed description of the library
                and how to order individual BAC clones, the entire library, or
                filters for hybridization from the BACPAC Resource Center can be
                found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
    source
        1..843
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="RPCI-98"
            /clone="BACR26H19"
            /note="end : TET3"

BASE COUNT      131 a 102 c 259 g 141 t 210 others
ORIGIN

Query Match      1.8%; Score 59.6; DB 122; Length 843;
Best Local Similarity 33.9%; Pred. No. 0.00011;
Matches 115; Conservative 67; Mismatches 157; Indels 0; Gaps 0;

QY 1 taaacacgaaggttatctctcaatcatctggttataataataatttttcccttng 60
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 829 WAAWATAATATATTTTAAAAATTTTDTTWDAAATAATTTATATATATTTTAT 770

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 10:24:43 ; Search time 6962.07 Seconds
(without alignments)
142.272 Million cell updates/sec

Title: US-09-183-972-1
Perfect score: 555
Sequence: 1 atttttttcgaacggggt.....agctgcgtactcagtcagta 555

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 972840 seqs, 892348106 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb_bal:*
- 2: gb_bal2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pil:*
- 8: gb_pil2:*
- 9: gb_prl:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sts:*
- 14: gb_sy:*
- 15: gb_un:*
- 16: em_fun:*
- 17: em_hum1:*
- 18: em_hum2:*
- 19: em_in:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_p1:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_sy:*
- 29: em_un:*
- 30: em_vi:*
- 31: gb_htg1:*
- 32: gb_htg2:*
- 33: gb_in1:*
- 34: gb_in2:*
- 35: em_bal:*
- 36: em_bal2:*
- 37: em_hum3:*
- 38: em_hum4:*
- 39: gb_pr4:*
- 40: gb_htg3:*
- 41: gb_htg4:*
- 42: gb_htg5:*
- 43: gb_htg6:*

- 44: gb_htg7:*
- 45: em_htg1:*
- 46: em_htg2:*
- 47: em_htg3:*
- 48: em_hum5:*
- 49: gb_p13:*
- 50: gb_pr5:*
- 51: gb_htg8:*
- 52: gb_htg9:*
- 53: gb_htg10:*
- 54: gb_htg11:*
- 55: gb_htg12:*
- 56: gb_htg13:*
- 57: gb_htg14:*
- 58: gb_in3:*
- 59: gb_htg15:*
- 60: gb_htg16:*
- 61: gb_htg17:*
- 62: em_htg4:*
- 63: em_htg5:*
- 64: em_htg6:*
- 65: em_htg7:*
- 66: em_hum6:*
- 67: gb_htg18:*
- 68: gb_htg19:*
- 69: gb_htg20:*
- 70: gb_htg21:*
- 71: gb_htg22:*
- 72: gb_htg23:*
- 73: gb_htg24:*
- 74: gb_htg25:*
- 75: gb_htg26:*
- 76: gb_htg27:*
- 77: gb_htg28:*
- 78: gb_htg29:*
- 79: gb_htg30:*
- 80: gb_htg31:*
- 81: gb_vil:*
- 82: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	555	100.0	555	11	AF047491	Macaca fa	
2	112	20.2	3268	39	AF047492	AF047492 Homo sapi	
3	85	15.3	422	11	HSIMPG03	AF017762 Homo sapi	
c	4	85	15.3	194704	32	AL157379	AL157379 Homo sapi
	5	43	7.7	537	11	HSIMPG02	AF017761 Homo sapi
6	41	7.4	311	11	HSIMPG06	AF017765 Homo sapi	
7	41	7.4	194704	32	AL157379	AL157379 Homo sapi	
8	30	5.4	300	11	HSIMPG04	AF017763 Homo sapi	
9	24	4.3	377	11	HSIMPG07	AF017766 Homo sapi	
10	23	4.1	304	11	HSIMPG05	AF017764 Homo sapi	
11	22	4.0	170439	42	AC013277	AC013277 Homo sapi	
12	22	4.0	183375	55	AC017079	AC017079 Homo sapi	
13	21	3.8	164180	68	AC009837	AC009837 Homo sapi	
14	21	3.8	169913	72	AC031980	AC031980 Homo sapi	
15	20	3.6	132805	10	HS339A18	297054 Human DNA s	
16	20	3.6	191079	32	HS1036D20	AL109851 Homo sapi	
17	19	3.4	758	13	AF106667	AF106667 Mus muscu	
18	19	3.4	49386	54	AC021257	AC021257 Homo sapi	
19	19	3.4	60787	78	AC061964	AC061964 Homo sapi	
20	19	3.4	73666	11	HSDJ581P3	AL109805 Human DNA	
21	19	3.4	116459	39	AF222684	AF222684 Homo sapi	
22	19	3.4	126579	39	AC004692	AC004692 Homo sapi	
c	23	19	3.4	151432	74	AC021544	AC021544 Homo sapi
	24	19	3.4	153858	31	AP001811	AP001811 Homo sapi

c 25 19 3.4 158230 56 AC012134 Homo sapi
 26 19 3.4 159946 73 AC015671 Homo sapi
 27 19 3.4 161591 68 AC024711 Homo sapi
 28 19 3.4 164017 11 AC005304 Homo sapi
 c 29 19 3.4 165772 68 AC025526 Homo sapi
 c 30 19 3.4 167677 31 AF001158 Homo sapi
 c 31 19 3.4 170398 32 CENS01DTH Homo sapi
 c 32 19 3.4 176126 31 AF000762 Homo sapi
 c 33 19 3.4 196074 56 AC024442 Homo sapi
 c 34 19 3.4 196414 68 AC025413 Homo sapi
 c 35 19 3.4 201744 80 AC055758 Homo sapi
 c 36 19 3.4 206623 54 AC023248 Homo sapi
 c 37 19 3.4 207571 31 AF000844 Homo sapi
 c 38 19 3.4 212411 31 AF001159 Homo sapi
 c 39 19 3.4 213856 31 AF001321 Homo sapi
 c 40 18 3.2 425 13 G31950 Chick
 c 41 18 3.2 431 13 G31948 Chick
 c 42 18 3.2 1713 4 GGHOMXR Homo sapi
 c 43 18 3.2 1901 12 MW074079 Mus muscu
 c 44 18 3.2 2111 8 AF071862 Vigna ung
 c 45 18 3.2 2713 12 AB03030859 Mus muscu

ALIGNMENTS

Credit date 10/2/00
 24-FEB-1998

RESULT 1
 AF047491 555 bp mRNA PRI
 LOCUS Macaca fascicularis interphotoreceptor matrix proteoglycan 150
 DEFINITION mRNA, partial cds.
 ACCESSION AF047491
 VERSION AF047491.1 GI:2906229
 KEYWORDS crab-eating macaque.
 SOURCE Macaca fascicularis
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
 Macaca.
 REFERENCE 1 (bases 1 to 555)
 Kuehn, M.H. and Hageman, G.S.
 Characterization And Complete cDNA Sequence Of IPM 150, A Novel
 Human Photoreceptor Cell-Associated Chondroitin-Sulfate
 Proteoglycan
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 555)
 Kuehn, M.H. and Hageman, G.S.
 Direct Submission
 TITLE Submitted (10-FEB-1998) Ophthalmology and Visual Sciences,
 University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA
 JOURNAL 52240, USA

FEATURES
 source
 1. .555
 /organism="Macaca fascicularis"
 /db_xref="taxon:9541"
 /tissue_type="retina"
 <1..>555
 /note="IPM 150"
 /codon_start=1
 /product="interphotoreceptor matrix proteoglycan 150"
 /protein_id="AAC03788.1"
 /db_xref="GI:2906230"
 /translation="IFPPNGVKVCPQESMKQILASQAYRLRVCOEAVWEAYRIFLD
 RIPDGEYDQWVSFCQETFCFLDIGNFNSQOEHLDLQQRKQSFPERKDEVSTE
 KTLGPSTVIVSTVAVSLGVPFPVPTDPTLNEILDNALNDRKPTTRETRELAYS
 EQRVLSLSLINORFKAELADQS"
 BASE COUNT 162 a 131 c 131 g 131 t
 ORIGIN

Query Match 100.0%; Score 555; DB 11; Length 555;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attttttccgaacgggggttaagtctgtccacaggaatccatgaacagatttttagcc 60
 |||||
 Db 1 ATTTTCTTTCCGACGCGGGTTAAAGTCTGTCCACAGGAATCCATGAACAGATTTAGCC 60
 |||||
 QY 61 agtttcaagcttattatagattgagagtggtgtcaggaagcagttatgggaagcattatcg 120
 |||||
 Db 61 AGTTTCAAGCTTATTATAGATTGAGAGTGTGTGAGGAAGCAGTATGGGAAGCATATCGG 120
 |||||
 QY 121 atctttcgaatcgcctccctgcacacagggggaatatcagagctgggtcagcttctgcca 180
 |||||
 Db 121 ATCTTTCTGATCGATCCATCCCTTGACACAGGGGAATATCAGGACTGGGTGAGCTTCTGCCAG 180
 |||||
 QY 181 cagagagcctttctcctcttttgacatcggaacaaacttcagcaattcccccagagcacctg 240
 |||||
 Db 181 CAGGAGACCTTCTGCTCTTTTGACATCGGACAAACTTCAGCAATTCCTCCAGGACACCTG 240
 |||||
 QY 241 gatcttccagcagagaataaaacagagaagtttccctcgagagaaaagatgaagtatct 300
 |||||
 Db 241 GATCTTCTCCAGCAGAGATAAAACAGAGAAGTTTCCCTGAGAGAAAAGATGAAGTATCT 300
 |||||
 QY 301 acagagaagacattggagagccttagtgaacacattgtggtttcaacagatgttgcacg 360
 |||||
 Db 301 ACAGAGAAGACATTGGGAGAGCCTTAGTGAACACCATTTGTGGTTTCAACAGATGTTGCCAGC 360
 |||||
 QY 361 gtctcacttgggcttccctgtcactcctgatgacacccctcctcaatgaattctcgat 420
 |||||
 Db 361 GTCTCACTTGGGCTCTTCCCTGTCTACTCTCTGATGACACCTCTCTCAATGAATTTCTCGAT 420
 |||||
 QY 421 aatgcactcaacgacacacaaagatcctacaacagagaagaaacagaactcgtgtgtct 480
 |||||
 Db 421 AATGCACCTCAACGACACACCAAGATGCTTACAACAGAGAAGAAACAGAACTCGCTGTGTCT 480
 |||||
 QY 481 gagagagagaggtggagctcagcatctctctgtataaacacagaggttcaagcagagctc 540
 |||||
 Db 481 GAGAGAGAGAGGTGGAGCTCAGCATCTCTCTGATAAACACAGAGTTTCAAGGCAGAGCTC 540
 |||||
 QY 541 gctgactctcagtc 555
 |||||
 Db 541 GCTGACTCTCAGTCA 555
 |||||
 RESULT 2
 AF047492 3268 bp mRNA PRI 26-OCT-1999
 LOCUS Homo sapiens interphotoreceptor matrix proteoglycan 150 (IMP1)
 DEFINITION mRNA, complete cds.
 ACCESSION AF047492
 VERSION AF047492.2 GI:6118565
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 3268)
 Kuehn, M.H. and Hageman, G.S.
 AUTHORS Expression and characterization of the IPM 150 gene (IMP1)
 TITLE product, a novel human photoreceptor cell-associated
 chondroitin-sulfate proteoglycan
 JOURNAL Matrix Biol. 18 (5), 509-518 (1999)
 MEDLINE 20068045
 REFERENCE 2 (bases 1 to 3268)
 Kuehn, M.H. and Hageman, G.S.
 AUTHORS Direct Submission
 TITLE Submitted (10-FEB-1998) Ophthalmology and Visual Sciences,
 University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA
 JOURNAL 52240, USA
 REFERENCE 3 (bases 1 to 3268)
 Kuehn, M.H. and Hageman, G.S.
 AUTHORS Direct Submission
 TITLE Submitted (26-OCT-1999) Ophthalmology and Visual Sciences,
 University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA
 JOURNAL 52240, USA

REMARK Sequence update by submitter
COMMENT On Oct 26, 1999 this sequence version replaced gi:2906231.
FEATURES Location/Qualifiers
source 1. .3268
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="between markers D6S280 and D6S1659"
/tissue_type="retina"
gene 1. .3268
/gene="IMP1"
CDS 131..2524
/gene="IMP1"
/codon_start=1
/note="IPM 150"
/product="interphotoreceptor matrix proteoglycan 150"
/protein_id="AAC03789.2"
/db_xref="GI:6118566"
/translation="MYLETRAIFFVFIQVQTKDISINIIYHSEYKIDIDPPRNET
TESTEKMYKMTWRIRIFDLAKHRTKRSFAFFTGKVCQESMKQIILSLQAYLRVC
QEAWEAYRIFLRIPDTGEYQDWVSIQQQETFCFLDGKFNFSQOEHLDLQORIKQ
RSPDRETSIAKTLGEGETIVISDVANVSLGPFPLTPDDTLNELLNDLNDIK
MPITERTEFAVLEEQVELSVLVNOKFAELADSSPYQELAGSLOLOKIFKK
LPGEKHLVLFGRPKKEDSSSTEMQLTAIFKRHSAEAKSPADLLSFDNSKIESE
VINGTEEDKPEIILATDLKRLISRALEEQSLDVGTIQFTDEIRGSLPAFGDIO
SELPTFAVITEDATISPELPEPQLETVDAEHLGPDTSWSPAMASTLSLSPFP
FMASISFLVTDQGTDTMATDQTMVPLGTIPTSDYSAISQALGISHPPASSDPSRS
SAGGEDMVRLEMDLSDTAPASEVSELSVYPDHFLEDTTPVSALQYLTTSMTI
APKRELVFSLRANAFNSDLFNKSSLEYRALEQOFTLLVPLKNSLTGFKOLE
ILNFRGSLVNSKMFSAKVPYNLTAKVGVLEDFRMAAQHLHLEDSYSLIEPA
LDQDCKFLACGEFAQCCKNTEEEACRCRPGDGSGLDGLFPGLCGPGTKECEVL
QGGAPCRLPDHSENQAYKTSVKRFQNNKVISKRNSLELLTVEYEEFNQDWEGN"
BASE COUNT 1044 a 724 c 564 g 836 t
ORIGIN

Query Match 20.28; Score 112; DB 39; Length 3268;
Best Local Similarity 100.0%; Pred. No. 1.6e-52;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 cagcttcagcttattatagattgagagtggtgtcaggagcagtgatgggaagcattatcg 119
|||||
DB 403 CAGCTTCAAGCTTATTATAGATTGAGAGTGTCAGGAAGCAGTATGGGAGCATATCG 462
|||||

QY 120 gatcttctgagtcacccctgcacaggggaatatcaggactgggtcagc 171
|||||
DB 463 GATCTTCTGATCGCATCCCTGCACAGGGGATATCAGGACTGGGTGAGC 514
|||||

RESULT 3
HSIMPG03
LOCUS HSIMPG03 422 bp DNA PRI 28-OCT-1998
DEFINITION Homo sapiens interphotoreceptor matrix gene (IPM150), exon 3.
ACCESSION AF011762
VERSION AF011762.1 GI:3800717
KEYWORDS 3 of 17
SEGMENT
SOURCE human sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 422)
AUTHORS Felbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
Schmid, M. and Weber, B.H.
TITLE Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMP1) gene: a candidate
for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 422)
AUTHORS Gehrig, A., Felbor, U., Kiesel, R., Hunt, D.M., Maunee-Hussels, I.E.
and Weber, B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)

localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choroideremia atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRI)
Unpublished
JOURNAL 3 (bases 1 to 422)
REFERENCE
AUTHORS Felbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES Location/Qualifiers
source 1. .422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
112..278
/gene="IPM150"
BASE COUNT 113 a 90 c 103 g 116 t
ORIGIN

Query Match 15.3%; Score 85; DB 11; Length 422;
Best Local Similarity 100.0%; Pred. No. 4.2e-37;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 agtgttcaggagcagtgatgggaagcattatggatcttttggatcgatccctgacac 146
|||||
DB 110 AGTGTGTCAGGAGCAGTATGGAGCATATCGGATCTTTCTGGATCGCATCCCTGCAC 169
|||||

QY 147 aggggaataatcaggactgggtcagc 171
|||||
DB 170 AGGGGAATATCAGGACTGGGTGAGC 194
|||||

RESULT 4
AL157379/c
LOCUS AL157379 194704 bp DNA HTG 22-MAR-2000
DEFINITION Homo sapiens chromosome 6 clone RP1-62L18, *** SEQUENCING IN
PROGRESS ***, 28 unordered pieces.
ACCESSION AL157379
VERSION AL157379.2 GI:7159486
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194704)
AUTHORS Sims, S.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Mar 6, 2000 this sequence version replaced gi:7009540.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00025 Length: 6617bp
Contig_ID: 00126 Length: 847bp
Contig_ID: 00151 Length: 1873bp
Contig_ID: 00275 Length: 6849bp
Contig_ID: 00307 Length: 1124bp
Contig_ID: 00337 Length: 3565bp
Contig_ID: 00377 Length: 5047bp
Contig_ID: 00383 Length: 1165bp
Contig_ID: 00410 Length: 286bp
Contig_ID: 00414 Length: 15535bp
Contig_ID: 00462 Length: 19955bp
Contig_ID: 00501 Length: 1028bp
Contig_ID: 00598 Length: 7288bp
Contig_ID: 00686 Length: 6841bp

Contig_ID: 00766 Length: 1115bp
Contig_ID: 00821 Length: 12763bp
Contig_ID: 00909 Length: 23204bp
Contig_ID: 00911 Length: 7481bp
Contig_ID: 00925 Length: 6805bp
Contig_ID: 00932 Length: 1460bp
Contig_ID: 00934 Length: 10010bp
Contig_ID: 01026 Length: 2716bp
Contig_ID: 01069 Length: 2283bp
Contig_ID: 01121 Length: 10059bp
Contig_ID: 01185 Length: 2031bp
Contig_ID: 01186 Length: 1779bp
Contig_ID: 01370 Length: 1640bp
Contig_ID: 01386 Length: 1498bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6617: contig of 6617 bp in length
* 6618 7417: gap of 800 bp
* 7418 15864: contig of 8447 bp in length
* 15865 16664: gap of 800 bp
* 16665 18537: contig of 1873 bp in length
* 18538 19337: gap of 800 bp
* 19338 26186: contig of 6849 bp in length
* 26187 26986: gap of 800 bp
* 26987 28110: contig of 1124 bp in length
* 28111 28910: gap of 800 bp
* 28911 32475: contig of 3565 bp in length
* 32476 33275: gap of 800 bp
* 33276 38322: contig of 5047 bp in length
* 38323 39122: gap of 800 bp
* 39123 40287: contig of 1165 bp in length
* 40288 41087: gap of 800 bp
* 41088 43973: contig of 2886 bp in length
* 43974 44773: gap of 800 bp
* 44774 60308: contig of 15535 bp in length
* 60309 61108: gap of 800 bp
* 61109 81103: contig of 19995 bp in length
* 81104 81903: gap of 800 bp
* 81904 82931: contig of 1028 bp in length
* 82932 83731: gap of 800 bp
* 83732 91019: contig of 7288 bp in length
* 91020 91819: gap of 800 bp
* 91820 98660: contig of 6841 bp in length
* 98661 99460: gap of 800 bp
* 99461 100575: contig of 1115 bp in length
* 100576 101375: gap of 800 bp
* 101376 114138: contig of 12763 bp in length
* 114139 114938: gap of 800 bp
* 114939 138142: contig of 23204 bp in length
* 138143 138942: gap of 800 bp
* 138943 146423: contig of 7481 bp in length
* 146424 147223: gap of 800 bp
* 147224 154028: contig of 6805 bp in length
* 154029 154828: gap of 800 bp
* 154829 156288: contig of 1460 bp in length
* 156289 157088: gap of 800 bp
* 157089 167098: contig of 10010 bp in length
* 167099 167898: gap of 800 bp
* 167899 170614: contig of 2716 bp in length
* 170615 171414: gap of 800 bp
* 171415 173637: contig of 2283 bp in length
* 173638 174497: gap of 800 bp
* 174498 184556: contig of 10059 bp in length
* 184557 185356: gap of 800 bp
* 185357 187387: contig of 2031 bp in length
* 187388 188187: gap of 800 bp
* 188188 189966: contig of 1779 bp in length

* 189967 190766: gap of 800 bp
* 190767 192406: contig of 1640 bp in length
* 192407 193206: gap of 800 bp
* 193207 194704: contig of 1498 bp in length.
FEATURES
source
Location/Qualifiers
1. .194704
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP1-62L18"
/clone_lib="RPC1-1"
BASE COUNT 53528 a 32672 c 33519 g 53383 t 21602 others
ORIGIN
Query Match 15.3% Score 85; DB 32; Length 194704;
Best Local Similarity 100.0%; Pred. No. 3.4e-37;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 87 agtgtgcaggagcagtcggaagcagcatcgcggtcttctggtatcgatccctgacac 146
|||||
Db 124770 AGTGTGTCAGGAACGATGCGGAAGCATATCGATCTTCTGGATCGATCCCTGACAC 124711
|||||
QY 147 aggggaatatcaggactgggtcagc 171
|||||
Db 124710 AGGGAATATCAGGACTGGGTGACG 124686
|||||
RESULT 5
HSIMPG02 537 bp DNA PRI 28-OCT-1998
LOCUS HSIMPG02 Homo sapiens interphotoreceptor matrix gene (IPM150), exon 2.
DEFINITION AF017761
ACCESSION AF017761
VERSION AF017761.1 GI:3800716
KEYWORDS 2 of 17
SEGMENT human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 537)
AUTHORS Felbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
Schmid, M. and Weber, B.H.
Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMP1) gene: a candidate
for 6q-linked retinopathies
Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
JOURNAL 98358139
MEDLINE 2 (bases 1 to 537)
REFERENCE Gehrig, A., Felbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussells, I.E.
and Weber, B.H.F.
Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choroidretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRL)
Unpublished
JOURNAL 3 (bases 1 to 537)
REFERENCE Felbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.
AUTHORS Direct Submission
TITLE Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
JOURNAL Hubland, Wuerzburg D-97074, Germany
FEATURES Location/Qualifiers
source
1. .537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
97. .330
/gene="IPM150"
BASE COUNT 195 a 99 c 74 g 169 t
ORIGIN
exon

```

Query Match          7.7%; Score 43; DB 11; Length 537;
Best Local Similarity 100.0%; Pred. No. 3.6e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 ggggttaagtctgcacaggaatccatgaacagattttag 58
|||||
Db 258 GGGGTAAAGTCTGTCACAGGAATCCATGAACAGATTITAG 300
|||||

RESULT 6
HSIMPG06          311 bp      DNA      PRI      28-OCT-1998
LOCUS      Homo sapiens interphotoreceptor matrix gene (IPM150), exon 6.
DEFINITION      AF017765
ACCESSION      AF017765
VERSION      AF017765.1 GI:3800720
KEYWORDS
SEGMENT
SOURCE
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 311)
AUTHORS      Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
Schmid,M. and Weber,B.H.
TITLE      Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate
for 6q-linked retinopathies
JOURNAL      Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE      98358139
REFERENCE      2 (bases 1 to 311)
AUTHORS      Gehrig,A., Felbor,U., Kelsell,R., Hunt,D.M., Maumenee-Hussels,I.B.
and Weber,B.H.F.
TITLE      Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal chorioretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRI)
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 311)
AUTHORS      Felbor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.
TITLE      Direct Submission
JOURNAL      Submitted (09-AUG-1997) Humanogenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES
source
1..311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
139..242
/gene="IPM150"
BASE COUNT      95 a 70 c 45 g 101 t
ORIGIN

Query Match          7.4%; Score 41; DB 11; Length 311;
Best Local Similarity 100.0%; Pred. No. 5e-12;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 tcaactctgatgacacctctcaatgaattcgataat 423
|||||
Db 172 TCACCTCTGATGACACCCCTCCTCAATGAATTCGATAAT 212
|||||

RESULT 7
AL157379          194704 bp      DNA      HTG      22-MAR-2000
LOCUS      Homo sapiens chromosome 6 clone RPL-62L18, *** SEQUENCING IN
DEFINITION      PROGRESS ***, 28 unordered pieces.
ACCESSION      AL157379
VERSION      AL157379.2 GI:7159486
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 194704)
Sims,S.
Direct Submission
Submitted (22-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Mar 6, 2000 this sequence version replaced gi:7009540.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00025 Length: 6617bp
Contig_ID: 00126 Length: 8447bp
Contig_ID: 00151 Length: 1873bp
Contig_ID: 00275 Length: 6849bp
Contig_ID: 00307 Length: 1124bp
Contig_ID: 00337 Length: 3565bp
Contig_ID: 00377 Length: 5047bp
Contig_ID: 00383 Length: 1165bp
Contig_ID: 00410 Length: 2886bp
Contig_ID: 00414 Length: 1535bp
Contig_ID: 00462 Length: 1995bp
Contig_ID: 00501 Length: 1028bp
Contig_ID: 00598 Length: 7288bp
Contig_ID: 00686 Length: 6841bp
Contig_ID: 00766 Length: 1115bp
Contig_ID: 00821 Length: 12763bp
Contig_ID: 00909 Length: 23204bp
Contig_ID: 00911 Length: 7481bp
Contig_ID: 00925 Length: 6805bp
Contig_ID: 00932 Length: 1460bp
Contig_ID: 00934 Length: 10010bp
Contig_ID: 01026 Length: 2716bp
Contig_ID: 01069 Length: 2283bp
Contig_ID: 01121 Length: 10059bp
Contig_ID: 01185 Length: 2031bp
Contig_ID: 01186 Length: 1779bp
Contig_ID: 01370 Length: 1640bp
Contig_ID: 01386 Length: 1498bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 6617: contig of 6617 bp in length
* 6618 7417: gap of 800 bp
* 7418 15864: contig of 8447 bp in length
* 15865 16664: gap of 800 bp
* 16665 18537: contig of 1873 bp in length
* 18538 19337: gap of 800 bp
* 19338 26186: contig of 6849 bp in length
* 26187 26986: gap of 800 bp
* 26987 28110: contig of 1124 bp in length
* 28111 28910: gap of 800 bp
* 28911 32475: contig of 3565 bp in length
* 32476 33275: gap of 800 bp
* 33276 38322: contig of 5047 bp in length
* 38323 39122: gap of 800 bp
* 39123 40287: contig of 1165 bp in length
* 40288 41087: gap of 800 bp
* 41088 43973: contig of 2886 bp in length
* 43974 44773: gap of 800 bp
* 44774 60308: contig of 15535 bp in length
* 60309 61108: gap of 800 bp
* 61109 81103: contig of 19995 bp in length
* 81104 81903: gap of 800 bp

```

```

* 81904 82931: contig of 1028 bp in length
* 82932 83731: gap of 800 bp
* 83732 91019: contig of 7288 bp in length
* 91020 91819: gap of 800 bp
* 91820 98660: contig of 6841 bp in length
* 98661 99460: gap of 800 bp
* 99461 100575: contig of 1115 bp in length
* 100576 101375: gap of 800 bp
* 101376 114138: contig of 12763 bp in length
* 114139 114938: gap of 800 bp
* 114939 138142: contig of 23204 bp in length
* 138143 138942: gap of 800 bp
* 138943 146433: contig of 7481 bp in length
* 146424 147223: gap of 800 bp
* 147224 154028: contig of 6805 bp in length
* 154029 154828: gap of 800 bp
* 154829 156288: contig of 1460 bp in length
* 156289 157088: gap of 800 bp
* 157089 167098: contig of 10010 bp in length
* 167099 167898: gap of 800 bp
* 167899 170614: contig of 2716 bp in length
* 170615 171414: gap of 800 bp
* 171415 173697: contig of 2283 bp in length
* 173698 174497: gap of 800 bp
* 174498 184556: contig of 10059 bp in length
* 184557 185356: gap of 800 bp
* 185357 187387: contig of 2031 bp in length
* 187388 188187: gap of 800 bp
* 188188 189966: contig of 1779 bp in length
* 189967 190766: gap of 800 bp
* 190767 192406: contig of 1640 bp in length
* 192407 193206: gap of 800 bp
* 193207 194704: contig of 1498 bp in length.
FEATURES
    source
        1. .194704
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone_lib="RPC1-1"
BASE COUNT 53528 a 32672 c 33519 g 53383 t 21602 others
ORIGIN
    Location/Qualifiers
        1. .194704
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone_lib="RPC1-1"

Query Match 7.4%; Score 41; DB 32; Length 194704;
Best Local Similarity 100.0%; Pred. No. 4.1e-12;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 tcactctgatgacacctctcctcaatgaatctcgataat 423
|||||
Db 149566 TCACCTCTGATGACACCTCTCAATGAATTCGATAAT 149606

RESULT 8
HSIMPG04 300 bp DNA PRI 28-OCT-1998
LOCUS Homo sapiens interphotoreceptor matrix gene (IPM150), exon 4.
DEFINITION AF017763
ACCESSION AF017763
VERSION AF017763.1 GI:3800718
KEYWORDS
SEGMENT 4 of 17
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 300)
AUTHORS Feibor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
Schmid,M., and Weber,B.H.
TITLE Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate
for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 377)
AUTHORS Gehrig,A., Feibor,U., Kelsell,R., Hunt,D.M., Maumenee-Hussells,I.E.
and Weber,B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRL)
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 377)
AUTHORS Feibor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES
    source
        1. .377
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="6q14.2-q15"

Query Match 5.4%; Score 30; DB 11; Length 300;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 gcagagaataaacagagaagtttcctga 281
|||||
Db 78 GCAGAGAATAAACAGAGAAGTTTCCTGA 107

RESULT 9
HSIMPG07 377 bp DNA PRI 28-OCT-1998
LOCUS Homo sapiens interphotoreceptor matrix gene (IPM150), exon 7.
DEFINITION AF017766
ACCESSION AF017766
VERSION AF017766.1 GI:3800721
KEYWORDS
SEGMENT 7 of 17
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 377)
AUTHORS Feibor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
Schmid,M., and Weber,B.H.
TITLE Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate
for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 377)
AUTHORS Gehrig,A., Feibor,U., Kelsell,R., Hunt,D.M., Maumenee-Hussells,I.E.
and Weber,B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRL)
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 377)
AUTHORS Feibor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES
    source
        1. .377
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="6q14.2-q15"

Query Match 5.4%; Score 30; DB 11; Length 300;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 gcagagaataaacagagaagtttcctga 281
|||||
Db 78 GCAGAGAATAAACAGAGAAGTTTCCTGA 107

FEATURES
    source
        1. .300
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="6q14.2-q15"
        82. .110
            /gene="IPM150"
BASE COUNT 108 a 43 c 54 g 95 t
ORIGIN
    Location/Qualifiers
        1. .300
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="6q14.2-q15"
        exon
            82. .110
            /gene="IPM150"

```

```

exon          163..303
              /gene="IPM150"
BASE COUNT    110 a      86 c      83 g      98 t
ORIGIN

Query Match   4.3%; Score 24; DB 11; Length 377;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 481 gagagcagaggggtggagctcagc 504
|||||
Db 190 GAGAGCAGAGGGTGGAGCTCAGC 213

RESULT 10
LOCUS          HSIMPG05      304 bp      DNA      PRI      28-OCT-1998
DEFINITION     Homo sapiens interphotoreceptor matrix gene (IPM150), exon 5.
ACCESSION      AF017764
VERSION        AF017764.1 GI:3800719
KEYWORDS       5 of 17
SEGMENT        human.
SOURCE         Homo sapiens
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      Felber, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
                Schmid, M., and Weber, B.H.
                Genomic organization and chromosomal localization of the
                interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate
                for 6q-linked retinopathies
                Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
                98358139
REFERENCE      2 (bases 1 to 304)
                Gehrig, A., Felber, U., Kelsell, R., Hunt, D.M., Maunee-Hussels, I.E.
                and Weber, B.H.F.
                Assessment of a novel interphotoreceptor matrix gene (IPM150)
                localized to 6q14.2-q15 in autosomal dominant Stargardt-like
                macular dystrophy, progressive bifocal chorioretinal atrophy
                (PBCRA), and North Carolina macular dystrophy (MCDRI)
                Unpublished
                3 (bases 1 to 304)
                Felber, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.
                Direct Submission
                Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
                Hubland, Wuerzburg D-97074, Germany
                Location/Qualifiers
                1..304
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /map="6q14.2-q15"
                107..171
                /gene="IPM150"
BASE COUNT     98 a      54 c      38 g      114 t
ORIGIN

exon
Query Match    4.1%; Score 23; DB 11; Length 304;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 cagagaagacattgggagagcct 324
|||||
Db 124 CAGAGAAGACATTGGGAGAGCCT 146

RESULT 11
AC013277/c
LOCUS          AC013277.2 GI:6289221
DEFINITION     Homo sapiens clone NH0480N09, WORKING DRAFT SEQUENCE, 1 unordered
                Center code: WUGSC
                ----- Genome Center -----
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC

```

```

Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: R_NH0462M09
----- Summary Statistics -----
Sequencing vector: M13; 84%
Sequencing vector: plasmid; 16%
Chemistry: Dye-terminator Big Dye; 18% of reads
Chemistry: Dye-terminator Big Dye; 18% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 177386 bases at least Q40
Consensus quality: 178778 bases at least Q30
Consensus quality: 179583 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 182375; sum-of-contigs
Quality coverage: 9.02 in Q20 bases; agarose-fp
Quality coverage: 9.22 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1568: contig of 1568 bp in length
* 1569 1668: gap of unknown length
* 1669 2744: contig of 1076 bp in length
* 2745 2844: gap of unknown length
* 2845 3995: contig of 1151 bp in length
* 3996 4095: gap of unknown length
* 4096 5379: contig of 1284 bp in length
* 5380 6820: contig of 1341 bp in length
* 6821 6920: gap of unknown length
* 6921 14913: contig of 7993 bp in length
* 14914 15013: gap of unknown length
* 15014 25762: contig of 10749 bp in length
* 25763 25863: gap of unknown length
* 25863 38947: contig of 12985 bp in length
* 38948 38947: gap of unknown length
* 38948 71070: contig of 32123 bp in length
* 71071 71170: gap of unknown length
* 71171 110883: contig of 39713 bp in length
* 110884 110984: gap of unknown length
* 110984 183375: contig of 72392 bp in length.
*
* Location/Qualifiers
*   1..183375
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="2"
*     /clone="RP11-462M9"
*
* BASE COUNT  49008 a 44555 c 43725 g 45076 t 1011 others
* ORIGIN
*
* Query Match      4.0%; Score 22; DB 55; Length 183375;
* Best Local Similarity 100.0%; Pred. No. 0.27;
* Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* Qy 157 caggactgggtcagctctgcc 178
*      |||||||
* Db 159020 CAGGACTGGGTCTGCTGCC 158999
*
* RESULT 13
* AC009837
* LOCUS      AC009837 164180 bp DNA HTG 24-MAR-2000
* DEFINITION Homo sapiens chromosome 17 clone RP11-550K23 map 17, WORKING DRAFT
* SEQUENCE, 16 unordered pieces.
* ACCESSION AC009837
* VERSION AC009837.2 GI:7321653
* KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 164180)

Birken, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 17, clone RP11-550K23

Unpublished

2 (bases 1 to 164180)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., Jones, C., Kann, L., Karatas, A., Lehocsky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.-J. and Zody, M.

Direct Submission

Submitted (02-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 24, 2000 this sequence version replaced gi:5819142.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L844

Center clone name: 550_K_23

----- Summary Statistics

Sequencing vector: M13; M77815; 96% of reads

3.823720025923535Chemistry: Dye-terminator Big Dye; 95% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 145704 bases at least Q40

Consensus quality: 153620 bases at least Q30

Consensus quality: 157666 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 162680; sum-of-contigs

Quality coverage:

* NOTE: This is a 'working draft' sequence. It currently

* consists of 16 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1948: contig of 1948 bp in length

* 1949 2048: gap of 100 bp

* 2049 5064: contig of 3016 bp in length

* 5065 5164: gap of 100 bp

* 5165 7712: contig of 2548 bp in length

* 7713 7812: gap of 100 bp

* 7813 10232: contig of 2420 bp in length

* 10233 10332: gap of 100 bp

* 10333 15696: contig of 5364 bp in length

* 15697 15796: gap of 100 bp

* 15797 21366: contig of 5570 bp in length

* 21367 21466: gap of 100 bp

* 21467 27213: contig of 5747 bp in length

* 27214 27313: gap of 100 bp

*	27314	33063:	contig of 5750 bp	in length
*	33064	33163:	gap of 100 bp	
*	33164	41063:	contig of 7900 bp	in length
*	41064	41163:	gap of 100 bp	
*	41164	50071:	contig of 8908 bp	in length
*	50072	50171:	gap of 100 bp	
*	50172	60777:	contig of 10606 bp	in length
*	60778	60877:	gap of 100 bp	
*	60878	72745:	contig of 11868 bp	in length
*	72746	72845:	gap of 100 bp	
*	72846	82971:	contig of 10126 bp	in length
*	82972	83071:	gap of 100 bp	
*	83072	103187:	contig of 20116 bp	in length
*	103188	103287:	gap of 100 bp	
*	103288	131753:	contig of 28466 bp	in length
*	131754	131853:	gap of 100 bp	
*	131854	164180:	contig of 32327 bp	in length.

Query Match 3.8%; Score 21; DB 68; Length 164180;
Best Local Similarity 100.0%; Pred. No. 1;

RESULT 14
AC031980/c

LOCUS	AC031980	16913 bp	DNA	HTG	03-APR-2000
DEFINITION	Homo sapiens clone RP11-330M20, WORKING DRAFT SEQUENCE, 44 unordered pieces.				

* NOTE: This is a 'working draft' sequence. It currently consists of 44 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
* 1 1436: contig of 1436 bp in length
* 1437 1536: gap of 100 bp
* 1537 2781: contig of 1245 bp in length
* 2782 2881: gap of 100 bp
* 2882 4320: contig of 1439 bp in length
* 4321 4420: gap of 100 bp
* 4421 5739: contig of 1319 bp in length
* 5740 5839: gap of 100 bp
* 5840 6907: contig of 1068 bp in length
* 6908 7007: gap of 100 bp
* 7008 8248: contig of 1241 bp in length
* 8249 8348: gap of 100 bp
* 8349 9788: contig of 1440 bp in length
* 9789 9888: gap of 100 bp
* 9889 10995: contig of 1107 bp in length
* 10996 11095: gap of 100 bp
* 11096 13665: contig of 2570 bp in length
* 13666 13765: gap of 100 bp
* 13766 14829: contig of 1064 bp in length
* 14830 14929: gap of 100 bp
* 14930 16057: contig of 1128 bp in length
* 16058 16157: gap of 100 bp
* 16158 17567: contig of 1410 bp in length
* 17568 17667: gap of 100 bp
* 17668 19356: contig of 1689 bp in length
* 19357 19456: gap of 100 bp
* 19457 21798: contig of 2342 bp in length
* 21799 21898: gap of 100 bp
* 21899 23416: contig of 1518 bp in length
* 23417 23516: gap of 100 bp
* 23517 25529: contig of 2013 bp in length
* 25530 25629: gap of 100 bp
* 25630 28082: contig of 2453 bp in length
* 28083 28182: gap of 100 bp
* 28183 30748: contig of 2566 bp in length
* 30749 30848: gap of 100 bp
* 30849 32852: contig of 2004 bp in length
* 32853 32952: gap of 100 bp
* 32953 34569: contig of 1617 bp in length
* 34570 34669: gap of 100 bp
* 34670 36546: contig of 1877 bp in length
* 36547 36646: gap of 100 bp
* 36647 39296: contig of 2650 bp in length
* 39297 39396: gap of 100 bp
* 39397 43689: contig of 4293 bp in length
* 43690 43789: gap of 100 bp
* 43790 47256: contig of 3467 bp in length
* 47257 47356: gap of 100 bp
* 47357 51218: contig of 3862 bp in length
* 51219 51318: gap of 100 bp
* 51319 54648: contig of 3330 bp in length
* 54649 54748: gap of 100 bp
* 54749 58151: contig of 3403 bp in length
* 58152 58251: gap of 100 bp
* 58252 60546: contig of 2295 bp in length
* 60547 60646: gap of 100 bp
* 60647 64897: contig of 4251 bp in length
* 64898 64997: gap of 100 bp
* 64998 68132: contig of 3125 bp in length
* 68123 68222: gap of 100 bp
* 68223 72915: contig of 4693 bp in length
* 72916 73015: gap of 100 bp
* 73016 76800: contig of 3785 bp in length
* 76801 76900: gap of 100 bp
* 76901 80193: contig of 3293 bp in length
* 80194 80293: gap of 100 bp
* 80294 85518: contig of 5225 bp in length
* 85519 85618: gap of 100 bp
* 85619 89640: contig of 4022 bp in length
* 89641 89740: gap of 100 bp
* 89741 98209: contig of 8469 bp in length
* 98210 98309: gap of 100 bp
* 98310 107872: contig of 9563 bp in length
```

```
* 107873 107972: gap of 100 bp
* 107973 113669: contig of 5697 bp in length
* 113670 113769: gap of 100 bp
* 113770 121589: contig of 7820 bp in length
* 121590 121689: gap of 100 bp
* 121690 129185: contig of 7496 bp in length
* 129186 129285: gap of 100 bp
* 129286 139326: contig of 10041 bp in length
* 139327 139426: gap of 100 bp
* 139427 150444: contig of 11018 bp in length
* 150445 150544: gap of 100 bp
* 150545 158766: contig of 8222 bp in length
* 158767 158866: gap of 100 bp
* 158867 169913: contig of 11047 bp in length.
```

FEATURES

```
source
1. .169913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC"
1. .1436
/note="assembly_fragment"
1537. .2781
/note="assembly_fragment"
2882. .4320
/note="assembly_fragment"
4421. .5739
/note="assembly_fragment"
5840. .6907
/note="assembly_fragment"
7008. .8248
/note="assembly_fragment"
8349. .9788
/note="assembly_fragment"
9889. .10995
/note="assembly_fragment"
11096. .13665
/note="assembly_fragment"
13766. .14829
/note="assembly_fragment"
14930. .16057
/note="assembly_fragment"
16158. .17567
/note="assembly_fragment"
17668. .19356
/note="assembly_fragment"
19457. .21798
/note="assembly_fragment"
21899. .23416
/note="assembly_fragment"
23517. .25529
/note="assembly_fragment"
25630. .28082
/note="assembly_fragment"
28183. .30748
/note="assembly_fragment"
30849. .32852
```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Query Match 3.88; Score 21; DB 72; Length 169913;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 cttctgccagcagcagacctt 191

|||||

Db 144806 CTTCTGCCAGCAGGACCTT 144786

RESULT 15

HS339A18/c

LOCUS HS339A18 132805 bp DNA

DEFINITION Human DNA sequence from PAC 339A18 on chromosome xpll.2. Contains

K1AA0178 gene, similar to mitosis-specific chromosome segregation protein SMC1 of S.cerevisiae, DNA binding protein similar to

PRI

23-NOV-1999

```

ACCESSION      URE-B1, ESTs and STS.
VERSION        297054.1 GI:2370077
KEYWORDS       DNA binding protein; KIAA0178; smc1; Xp11.2.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 132805)
               Direct Submission
               Submitted (05-SEP-1997) Chromosome X Project Group
               (http://www.sanger.ac.uk/HGP/ChrX/) Sanger Centre, Hinxton,
               Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
               humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
               On Sep 9, 1997 this sequence version replaced gi:2213437.
               IMPORTANT: This sequence is the entire insert of clone 339A18.
               During sequence assembly data is compared from overlapping clones.
               Where differences are found these are annotated as variations
               together
               with a note of the overlapping clone name. Note that the variations
               annotated may not be found in the sequence submission corresponding
               to
               the overlapping clone as we submit sequences with only a small
               overlap
               as described above.
               This sequence was generated from part of bacterial clone contigs of
               human chromosome X, constructed by the Sanger Centre chromosome X
               mapping group. Further information can be found at
               http://www.sanger.ac.uk/HGP/ChrX/
               This sequence has been finished according to sequence map criteria
               as follows. An attempt is made to resolve all sequencing problems,
               such as compressions and repeats, but not necessarily within known
               annotated human repeat sequence elements (e.g. Alu). Where the
               sequence is ambiguous, there is an annotation using the 'unsure'
               feature key.
               The true left end of clone 339A18 is at 1 in this sequence. The
               true right end of clone 339A18 is at 132805.
               339A18 is from the library RPC13 constructed at the Roswell Park
               Cancer
               Institute by the group of Pieter de Jong.
               For further details see http://bacpac.med.buffalo.edu/.
FEATURES       source
               1..132805
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="X"
               /map="Xp11.1-Xp11.4"
               /clone="RP3-339A18"
               /clone_lib="RPC1-3"
               616..669
               /note="MIR repeat: matches 217..156 of consensus"
               673..902
               /note="AluX repeat: matches 294..56 of consensus;
               incomplete repeat"
               927..1053
               /note="MIR repeat: matches 162..34 of consensus"
               complement(join(<1185..1322,1418..1564,1666..1845,
               1944..2129,5234..5441,5593..5675,7952..8092,8186..8444,
               9091..9329,9423..9626,10948..11060,11171..11359,
               18682..>18769))
               /note="match D80000; similar to mitosis-specific
               chromosome; segregation protein SMc1 of S.cerevisiae"
               /codon_start=1
               /product="KIAA0178"
               /protein_id="CAB09784.1"
               /db_xref="GI:2370078"
               /db_xref="SPTREMBL:O14995"
               /translation="EIENFKSKGRIQICGPORFATIGPNSGKSLMDAISFVLGE
               KTSNLVRKLDLHGAPVGPAAFRVSMVSYSEGAEDRTFARVIGGSSEYKINN
               KVVOLHYSELEKILIKARFLVFCQAVESIAMKNPKEFTALFEISRSGLAQE
               YDKRKEMVAEDTQFNHKKNTAAERKAQKEADRYORLKDDEVVRAQVOLQL
               FLYHNEVEIKNLKASKNEIEKKKRDMDKVEDEKKEKELGKMKMREQQIQIEK
               IKEKDSLNQKRPQYIKAKENTSHIKIKLEAKKSLQNAOKHYKKRKGDMDELEKML
               SVEKARQFEERMEESQSGRODITLEENQYKKYHRLKEASRAATLAOLEKFNKD
               QKADQRLDLEERKKEVTEAKIKOKLREIENQKRIELEIITTSQSEEBQKKLEG
               ELTEFEVMAKRRIIDINKELNQMEIQGDADIRQESSRQQRKAETIKRILYPGSV
               YGLRIDLCQPTKKQYGIATVKILGKNMDAIIIVDSKTRGDCIQYIKQRQEPETFLPL
               DYLEVPTEKRLKGLKLVIVIRYEPPIKALQYACGNALVCNVDVARFIAGF
               GKHAKETVALDGTILFQKSGVLSGGADLKAKARRWDEKAYDKLKEKKERLTELKEQM
               GARKEAEULRQVQSQAHGLQMRUKYSQSDLEQTKRHLALNLQ"
               /note="FLAM_A repeat: matches 2..132 of consensus"
               2280..2410
               /note="FLAM_A repeat: matches 2..132 of consensus"
               2871..3169
               /note="AluX repeat: matches 300..2 of consensus"
               3252..3330
               /note="MER21B repeat: matches 790..511 of consensus"
               3534..3968
               /note="MER21B repeat: matches 459..6 of consensus"
               4220..4507
               /note="AluX repeat: matches 1..286 of consensus"
               5008..5031
               /note="12 copies of 2 mer 96 & conserved"
               5033..5209
               /note="MIR repeat: matches 20..216 of consensus"
               5967..6007
               /note="MIR repeat: matches 108..150 of consensus"
               6087..6182
               /note="AluJb repeat: matches 135..45 of consensus;
               incomplete repeat"
               6385..6518
               /note="AluSg repeat: matches 1..131 of consensus;
               incomplete repeat"
               6519..6688
               /note="AluX repeat: matches 39..206 of consensus;
               incomplete repeat"
               6690..6993
               /note="AluSg repeat: matches 1..303 of consensus"
               7008..7066
               /note="AluX repeat: matches 192..250 of consensus;
               incomplete repeat"
               7067..7299
               /note="AluJb repeat: matches 85..296 of consensus;
               incomplete repeat"
               7303..7467
               /note="MIR repeat: matches 84..261 of consensus"
               7501..7798
               /note="AluSg repeat: matches 1..300 of consensus"
               7815..7888
               /note="MER46 repeat: matches 1..72 of consensus"
               7852..7906
               /note="MER46 repeat: matches 136..84 of consensus"
               9986..10286
               /note="AluSg repeat: matches 1..303 of consensus"
               10288..10323
               /note="18 copies of 2 mer 81 & conserved"
               10514..10703
               /note="MER5A repeat: matches 189..3 of consensus"
               11602..11874
               /note="AluSx repeat: matches 7..300 of consensus"
               11875..12047
               /note="AluJb repeat: matches 134..302 of consensus;
               incomplete repeat"
               12211..12513
               /note="AluSg repeat: matches 1..303 of consensus"
               12587..12880
               /note="AluSc repeat: matches 293..1 of consensus"
               12882..13044
               /note="L1MB2 repeat: matches 752..914 of consensus"
               13052..13127
               /note="MER42c repeat: matches 904..980 of consensus"
               13146..13276
               /note="FLAM_C repeat: matches 131..1 of consensus"
               13461..13762
               /note="AluSp repeat: matches 2..303 of consensus"
               14357..14623

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2000, 19:17:50 ; Search time 66.18 Seconds
(without alignments)
193.817 Million cell updates/sec

Title: US-09-183-972-2

Perfect score: 950

Sequence: 1 IFPPNGVKVCPQESMRQILA.....ELSLINQRFKAELADSQS 185

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL12.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	950	100.0	185	6	O46616	046616 macaca fasc
2	885	93.2	771	4	O43686	043686 homo sapien
3	885	93.2	797	4	O95094	095094 homo sapien
4	297.5	31.3	1239	11	P70628	P70628 rattus norv
5	89.5	9.4	356	2	O06979	O06979 bacillus su
6	88	9.3	737	4	O9Y2L7	O9Y2L7 homo sapien
7	85.5	9.0	378	2	O85462	O85462 pasteurella
8	85.5	9.0	1520	4	O9Y4E0	O9Y4E0 homo sapien
9	85	8.9	226	2	O32269	O32269 bacillus su
10	84	8.8	5376	11	O88799	O88799 mus musculu
11	83	8.7	947	10	O04494	O04494 arabidopsis
12	82.5	8.7	1079	12	O9YW56	O9YW56 melanoplus
13	82	8.6	328	2	O3WYJ0	O3WYJ0 thermotoga
14	82	8.6	423	2	O34703	O34703 bacillus su
15	81	8.5	1197	4	O95347	O95347 homo sapien
16	81	8.5	4131	5	O19542	O19542 caenorhabdi
17	80	8.4	832	5	O16360	O16360 caenorhabdi
18	79.5	8.4	504	3	O14362	O14362 schizosacch
19	79.5	8.4	830	10	O04864	O04864 solanum tub

20	79.5	8.4	1083	4	O75146	O75146 homo sapien
21	79.5	8.4	2109	13	P79787	P79787 gallus gall
22	79	8.3	421	1	O28532	O28532 archaeoglob
23	79	8.3	2212	12	O39794	O39794 ebola virus
24	78.5	8.3	134	2	O86445	O86445 planktothri
25	78.5	8.3	134	2	O88099	O88099 planktothri
26	78.5	8.3	360	1	O58183	O58183 pyrococcus
27	78	8.2	362	2	O926L3	O926L3 chlamydia p
28	78	8.2	406	5	O17460	O17460 schistosoma
29	78	8.2	1047	5	O24019	O24019 drosophila
30	77	8.1	1545	5	O9XX31	O9XX31 caenorhabdi
31	76.5	8.1	341	2	O92IW2	O92IW2 salmonella
32	76.5	8.1	343	10	O82311	O82311 arabidopsis
33	76.5	8.1	523	4	O95404	O95404 homo sapien
34	76.5	8.1	1237	2	O84549	O84549 chlamydia t
35	76.5	8.1	3394	5	O77384	O77384 plasmodium
36	76	8.0	401	10	O92QL4	O92QL4 arabidopsis
37	76	8.0	900	10	O23454	O23454 arabidopsis
38	76	8.0	1017	5	O77371	O77371 plasmodium
39	76	8.0	1199	5	P91349	P91349 caenorhabdi
40	75.5	7.9	756	2	P72877	P72877 synecocyst
41	75.5	7.9	795	11	O92LS3	O92LS3 mus musculu
42	75.5	7.9	1099	2	O54377	O54377 lactococcus
43	75.5	7.9	1418	3	Q12267	Q12267 saccharomyc
44	75	7.9	500	2	P73537	P73537 synecocyst
45	75	7.9	649	5	O44633	O44633 caenorhabdi

ALIGNMENTS

RESULT 1

O46616 PRELIMINARY; PRT; 185 AA.

6-1-98

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

10269

RESULT 2
 ID 043686 PRELIMINARY; PRT; 771 AA.
 AC 043686;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE INTERPHOTORECEPTOR MATRIX PROTEOGLYCAN 150.
 GN IMPG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=RETINA;
 RP KUEHN M.H., HAGEMAN G.S.;
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF047492; AAC03789.1; -;
 DR PFAM; PF01390; SEA; 1.
 SQ SEQUENCE 771 AA; 86372 MW; 1580AF90 CRC32;

 Query Match 93.2%; Score 885; DB 4; Length 771;
 Best Local Similarity 90.8%; Pred. No. 8.4e-71;
 Matches 167; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

 QY 2 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 61
 DB 73 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 132
 QY 62 ETFCFLFDIGQNFNSQEHLDLQORIKORSPPERKDEVSCTEKTIGEPSETIVVSTDVAVS 121
 DB 133 ETFCFLFDIGQNFNSQEHLDLQORIKORSPPERKDEVSCTEKTIGEPSETIVVSTDVAVS 192
 QY 122 SLGFPFVTPDDTLNEILDNALNDTKMPTTERETELAVSEQRVELSISLQNRKFAELA 181
 DB 193 SLGFPFVTPDDTLNEILDNALNDTKMPTTERETELAVSEQRVELSISLQNRKFAELA 252
 QY 182 DSQS 185
 DB 253 DSQS 256

 RESULT 3
 ID 095094 PRELIMINARY; PRT; 797 AA.
 AC 095094;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE INTERPHOTORECEPTOR MATRIX.
 GN IPM150.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 98358139.
 RA FELBOR U., GERRIG A., SAUER C.G., MARQUARDT A., KOHLER M., SCHMID M.,
 RW WEBER B.H.;
 RT "Genomic organization and chromosomal localization of the
 RT 6q-linked retinopathies";
 RL Cytogenet. Cell Genet. 81:12-17(1998).
 DR EMBL; AF017766; AAC68835.1; JOINED.
 DR EMBL; AF017760; AAC68835.1; JOINED.
 DR EMBL; AF017761; AAC68835.1; JOINED.
 DR EMBL; AF017762; AAC68835.1; JOINED.
 DR EMBL; AF017763; AAC68835.1; JOINED.
 DR EMBL; AF017764; AAC68835.1; JOINED.
 DR EMBL; AF017765; AAC68835.1; JOINED.
 SQ SEQUENCE 797 AA; 89387 MW; 2369FDEA CRC32;

 Query Match 93.2%; Score 885; DB 4; Length 797;
 Best Local Similarity 90.8%; Pred. No. 8.8e-71;
 Matches 167; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

 QY 2 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 61
 DB 73 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 132
 QY 62 ETFCFLFDIGQNFNSQEHLDLQORIKORSPPERKDEVSCTEKTIGEPSETIVVSTDVAVS 121
 DB 133 ETFCFLFDIGQNFNSQEHLDLQORIKORSPPERKDEVSCTEKTIGEPSETIVVSTDVAVS 192
 QY 122 SLGFPFVTPDDTLNEILDNALNDTKMPTTERETELAVSEQRVELSISLQNRKFAELA 181
 DB 193 SLGFPFVTPDDTLNEILDNALNDTKMPTTERETELAVSEQRVELSISLQNRKFAELA 252
 QY 182 DSQS 185
 DB 253 DSQS 256

 RESULT 4
 ID P70628 PRELIMINARY; PRT; 1239 AA.
 AC P70628;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE PG10.2.
 GN PG10.2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 97038366.
 RA WANG X., BROWNSTEIN M.J., YOUNG W.S.;
 RT "Sequence analysis of PG10.2, a gene expressed in the pineal gland and
 RT the outer nuclear layer of the retina";
 RL Brain Res. Mol. Brain Res. 41:269-278(1996).
 DR EMBL; U76717; AAC52891.1; -;
 DR PFAM; PF01390; SEA; 2.
 KW Glycoprotein.
 SQ SEQUENCE 1239 AA; 137302 MW; 449D03AC CRC32;

 Query Match 31.3%; Score 297.5; DB 11; Length 1239;
 Best Local Similarity 33.3%; Pred. No. 2.7e-18;
 Matches 67; Conservative 39; Mismatches 62; Indels 33; Gaps 33;

 QY 1 IFFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 60
 DB 74 ILFPNGVRICPSDTVAEAVANVYFKARVCQEAIVEAFRTFDRLPGREYQYWMNLCE 133
 QY 61 QETFCFLFDIGQNFNSQEHLDLQORI---KQSFPERKQ-----EVSCTEKTIGEPSETI 112
 DB 134 DGVTSVFEMGTQFSQSEVHRHLMKEIKYTKAESSSSCKDAQCPULSPVPIGE----- 188
 QY 113 VVSTDVAVSISLGPFP-----VTPDDTLNEILDNALNDTKMPTTERETELAVSEQR 164

RESULT 2
 ID 043686 PRELIMINARY; PRT; 771 AA.
 AC 043686;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE INTERPHOTORECEPTOR MATRIX PROTEOGLYCAN 150.
 GN IMPG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=RETINA;
 RA KUEHN M.H., HAGEMAN G.S.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF047492; AAC03789.1; -;
 DR PFAM; PF01390; SEA; 1.
 SQ SEQUENCE 771 AA; 86372 MW; 1580AF90 CRC32;

 Query Match 93.2%; Score 885; DB 4; Length 771;
 Best Local Similarity 90.8%; Pred. No. 8.4e-71;
 Matches 167; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

 QY 2 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 61
 DB 73 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 132
 QY 62 ETFCFLDIGNFNSQEHLDLQRIKORSPPERKDEVSCTKTIGEPSETIVVSTDVASV 121
 DB 133 ETFCFLDIGNFNSQEHLDLQRIKORSPPERKDEVSCTKTIGEPSETIVVSTDVASV 192
 QY 122 SLGFPPTPDPTLLNEILDNLTNDTKMPTTERETELAVSEQRVELSVSLNQKFAELA 181
 DB 193 SLGFPPLTPDPTLLNEILDNLTNDTKMPTTERETEFVLEQRVELSVSLNQKFAELA 252
 QY 182 DSQS 185
 DB 253 DSQS 256

 RESULT 3
 ID 095094 PRELIMINARY; PRT; 797 AA.
 AC 095094;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE INTERPHOTORECEPTOR MATRIX.
 GN IPM150.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC MEDLINE; 98358139.
 RA FELBOR U., GERRIG A., SAUER C.G., MARQUARDT A., KOHLER M., SCHMID M.,
 RA WEBER B.H.;
 RT "Genomic organization and chromosomal localization of the
 RT 6q-linked retinopathies";
 RL Cytogenet. Cell Genet. 81:12-17(1998).
 DR EMBL; AF017766; AAC68835.1; JOINED.
 DR EMBL; AF017760; AAC68835.1; JOINED.
 DR EMBL; AF017761; AAC68835.1; JOINED.
 DR EMBL; AF017762; AAC68835.1; JOINED.
 DR EMBL; AF017763; AAC68835.1; JOINED.
 DR EMBL; AF017764; AAC68835.1; JOINED.
 DR EMBL; AF017765; AAC68835.1; JOINED.
 SQ SEQUENCE 797 AA; 89387 MW; 2369FDEA CRC32;

 Query Match 93.2%; Score 885; DB 4; Length 797;
 Best Local Similarity 90.8%; Pred. No. 8.8e-71;
 Matches 167; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

 QY 2 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 61
 DB 73 FFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 132
 QY 62 ETFCFLDIGNFNSQEHLDLQRIKORSPPERKDEVSCTKTIGEPSETIVVSTDVASV 121
 DB 133 ETFCFLDIGNFNSQEHLDLQRIKORSPPERKDEVSCTKTIGEPSETIVVSTDVASV 192
 QY 122 SLGFPPTPDPTLLNEILDNLTNDTKMPTTERETELAVSEQRVELSVSLNQKFAELA 181
 DB 193 SLGFPPLTPDPTLLNEILDNLTNDTKMPTTERETEFVLEQRVELSVSLNQKFAELA 252
 QY 182 DSQS 185
 DB 253 DSQS 256

 RESULT 4
 ID 095094 PRELIMINARY; PRT; 1239 AA.
 AC 095094;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE PG10.2.
 GN PG10.2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC MEDLINE; 97038366.
 RA WANG X., BROWNSTEIN M.J., YOUNG W.S.;
 RT "Sequence analysis of PG10.2, a gene expressed in the pineal gland and
 RT the outer nuclear layer of the retina";
 RL Brain Res. Mol. Brain Res. 41:269-278(1996).
 DR EMBL; U76717; AAC52891.1; -;
 DR PFAM; PF01390; SEA; 2.
 KW Glycoprotein.
 SQ SEQUENCE 1239 AA; 137302 MW; 449D03AC CRC32;

 Query Match 31.3%; Score 297.5; DB 11; Length 1239;
 Best Local Similarity 33.3%; Pred. No. 2.7e-18;
 Matches 67; Conservative 39; Mismatches 62; Indels 33; Gaps 33;

 QY 1 IFFPNGVKVCPQESMKQILASLQAYRLRVCOEAVWEAYRIFLDRIPTGETQDWNVSFCQ 60
 DB 74 ILFPNGVRICPSDTVAEAVANVYFKARVCOEAIWEAFRTFDRLPGREYQYWMNLCE 133
 QY 61 QETFCFLDIGNFNSQEHLDLQRIKORSPPERKDEVSCTKTIGEPSETIVVSTDVASV 112
 DB 134 DGVTSVFEMGTQFSQSEVHRHLMKEIKYTKAESSSSCKDAQCPULSPVPIGE----- 188
 QY 113 VVSTDVASVSLGPPF-----VTPDPTLLNEILDNLTNDTKMPTTERETELAVSEQR 164

```

Db 189 -TSLAGAVSASYPGAASERSAASPOESINEL-----ENVTEQTPP-----AASQI 236
QY 165 VELSILINORFKAEADSDS 185
Db 237 AEPISQLLGRQYSELDRPSS 257

RESULT 5
O06979
ID O06979 PRELIMINARY; PRT; 356 AA.
AC O06979;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE HYPOTHETICAL 40.2 KD PROTEIN.
GN YVQC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA DENIZOT F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRISSE R., BOURSIBER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRICH S.D., EMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAWATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTELELLA D., PORWOLLIK S., PRESCOTT A.M.,
RA PRESCHAN E., PUJIC P., PUKNELLE B., RAPPOORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHE E., ROCHE B., ROSE M., SADAIE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT "The complete genome sequence of the gram positive bacterium Bacillus
RT subtilis.";
RT Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z94043; CAB08063.1; -.
DR EMBL; Z99121; CAB15476.1; -.
DR PFAM; PF00512; signal; 1.
KW Hypothetical protein.
SQ SEQUENCE 356 AA; 40180 MW; 2940FA74 CRC32;

Query Match 9.4%; Score 89.5; DB 2; Length 356;
Best Local Similarity 25.7%; Pred. No. 1.9;

```

```

Matches 35; Conservative 26; Mismatches 40; Indels 35; Gaps 5;
QY 37 EAYRIFLDRIPDTGEYQDWVSCQOETFCFLDIGQNFNSQEHLDLQQRKQSFPERK 96
Db 172 ERFERGLDMQLQTARLEEFADVKQTFDLAEWVSLINQEK-----RQFIKRLFF--- 223
QY 97 DEVSTEKILGEPSETIVSTDVASVSLGFPFVTPDDTLLEILNDNALNDTKM-----PTTE 152
Db 224 -----TLHVPPNAVQISSDQKWLSE-----VVEQILFNALKYSKQGVGDPITI 266
QY 153 R-----ETELAVSEE 162
Db 267 RIETQGHETRLSVADE 282

RESULT 6
Q9Y2L7
ID Q9Y2L7 PRELIMINARY; PRT; 737 AA.
AC Q9Y2L7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE KIAA1014 PROTEIN (FRAGMENT).
GN KIAA1014.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 99245063.
RA NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., HIROSAWA M.,
RA MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023231; BAA76858.1; -.
FT NON_TER 1
SQ SEQUENCE 737 AA; 80758 MW; F71CA914 CRC32;

Query Match 9.3%; Score 88; DB 4; Length 737;
Best Local Similarity 26.1%; Pred. No. 6.2;
Matches 31; Conservative 20; Mismatches 50; Indels 18; Gaps 3;
QY 71 QNFSSQEH-----LDLLQQRKQSFPERKDEVSTKILGEPSETIVSTDVASV 121
Db 29 QALNSNEEEKKGVAASLLAPLPEGIEEERWRKVKICEEPVSEVKETSTTVEEATTI 88
QY 122 SLGPPFVTPDDTLLEILNDNALNDTKMPTTERETELAVSEORVELSILINQRFKAE 180
Db 89 -----VRPQEIIMLDNIEDPSQEDLCVSVQSGESE-----EEEEQDTLELVLKRAEL 138

RESULT 7
O85462
ID O85462 PRELIMINARY; PRT; 378 AA.
AC O85462;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE HEXC.
GN HEXC.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
RN [1]
RP SEQUENCE FROM N.A.
RA CHUNG J.Y., ZHANG Y., ADLER B.;
RT "The capsule biosynthetic locus of Pasteurella multocida A:1#.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067175; AAC67255.1; -.

```



```

Qy 98 EVSTEKILGPSETIVSTDVASVSLGFPVPTDDILLNEILNALNDKMPPTRETEL 157
    || | : : : || | : : || | : : || | : : || | : : || | : : ||
Db 447 ANSTAKDLSGKLGQKDEYEGDLNPDIQEDTV--EDSDSISNREIKNAEIKEA 504
    || | : : : || | : : || | : : || | : : || | : : || | : : ||
Qy 158 AVSEQRQVELSISLINORFAELADS 183
    | : : | : : | : | : | : | : | : | : | : | : | : | : | : |
Db 505 MVIKQRSDQESM-----QEEKSETRDS 526
    | : : | : : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12
Q9YW56 PRELIMINARY; PRT; 1079 AA.
AC Q9YW56
ID Q9YW56;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ORF MSV036 POTATIVE DNA POLYMERASE, CHORISTONEURA BIENNIS ENTOMOPOX
DE VIRUS DPOL HOMOLOG (VACCINIA E9L), SIMILAR TO SW:P30319.
GN MSY036.
OS Melanoplus sanguinipes entomopoxvirus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
RN [1]
   SEQUENCE FROM N.A.
   STRAIN-TUCSON;
   MEDLINE: 99102612.
RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
RL J. Virol. 73:533-552(1999).
RN [2]
   SEQUENCE FROM N.A.
   STRAIN-TUCSON;
RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;
RL Submitted (MAY 1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N
   PYROPHOSPHATE + DNA(N).
CC EMBL; AF063866; AAC97837.1; -
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW DNA-directed DNA polymerase; DNA replication; DNA-binding.
SQ SEQUENCE 1079 AA; 127406 MW; 3C26B08D CRC32;

```

Query Match	8.7%	Score 82.5;	DB 12;	Length 1079;
Best Local Similarity	18.1%	Pred. No. 30;		
Matches 35;	Conservative 41;	Mismatches 46;	Indels 71;	Gaps

Qy	21	SLQAYVRLVCOEAVWEAYRIFL-DRIPDTGEY-----QDWVSFCQETCFLD 68
		: : : : : : : : :
Db	834	SLNTYMLRICKD-----IGIFSTNTISNGTIPKLFNLEFGVGKKNMFLIEKKKVISYD 888
		: : : : : : : : : :
Qy	69	-----IGONFSNSOEH-----LDLQORIKORGFPERKDEVSTEKTL 105
		: : : : : : : : : :
Db	889	IINYDTLEAKIESKGTSLIKRDYSNFKHHKYKTIIETIIQKSIENKN----- 934
		: : : : : : : : : :
Qy	106	GEPSETIIVSTDVASVSGFPPTDPDTLLNEILDN--ALNDTKMPTTERETELAYSEEQ 163
		: : : : : : : : : : :
Db	935	-NNSKNVVIKY-----IDELIKNLVDNLKLYMDFAVTKRYSGKYSIDN 934
		: : : : : : : : : : :
Qy	164	RVELSISLINQRF 176
		: : : : : : : : : : :
Db	980	IIELTVNKFNAKY 992
		: : : : : : : : : : :

RESULT 13
Q9WYJ0 PRELIMINARY; PRT; 328 AA.
ID Q9WYJ0
AC Q9WYJ0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 37.6 KD PROTEIN.
OS TM0355.
GN Thermotoga maritima.
OS


```

DE CHROMOSOME-ASSOCIATED PROTEIN-E.
GN HCAP-E.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TERATOCARCINOMA;
RX MEDLINE: 99007239.
RA SCHMIESING J.A., BALL A.R. JR., GREGSON H.C., ALDERTON J.M., ZHOU S.,
RA YOKOMORI K.;
RT "Identification of two distinct human SMC protein complexes involved
RL in mitotic chromosome dynamics.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12906-12911(1998).
DR EMBL: AF092563; AAC72360.1; -.
SQ SEQUENCE 1197 AA; 135780 MW; 9E999CBF CRC32;

Query Match 8.5%; Score 81; DB 4; Length 1197;
Best Local Similarity 24.2%; Pred. No. 47;
Matches 23; Conservative 24; Mismatches 32; Indels 16; Gaps 3;

QY 71 QNFSNSQEHLLQORIKQSFPERKDEV-STKTLGEPSETIVVSTDVAVSLGFPVPT 129
DB 717 QWEMKTEADLLQTKLQSSYHKQEEELDKKKTIEESETLKNTKIQRKAEKY --- 773

QY 130 PDDTLLNEILDNALNDTKMPTTETELAVSEQR 164
DB 774 -----EVLEN-----KKNAEAEERELKDAQK 796

```

Search completed: September 3, 2000, 20:08:21
Job time: 3031 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 16:29:14 ; Search time 6962.07 Seconds
(without alignments)
1067.680 Million cell updates/sec

Title: US-09-183-972-5
Perfect score: 4165
Sequence: 1 cgggywayttgaaaggaca.....aaactactgttaaaaaaa 4165

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 972840 seqs, 892348106 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_om:
4: gb_ov:
5: gb_pat:
6: gb_ph:
7: gb_pl1:
8: gb_pl2:
9: gb_pr1:
10: gb_pr2:
11: gb_pr3:
12: gb_ro:
13: gb_sts:
14: gb_sy:
15: gb_un:
16: em_fun:
17: em_hum1:
18: em_hum2:
19: em_in:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sts:
28: em_sy:
29: em_un:
30: em_vi:
31: gb_htg1:
32: gb_htg2:
33: gb_in1:
34: gb_in2:
35: em_ba1:
36: em_ba2:
37: em_hum3:
38: em_hum4:
39: gb_pr4:
40: gb_htg3:
41: gb_htg4:
42: gb_htg5:
43: gb_htg6:

44: gb_htg7:
45: em_htg1:
46: em_htg2:
47: em_htg3:
48: em_hum5:
49: gb_pl3:
50: gb_pr5:
51: gb_htg8:
52: gb_htg9:
53: gb_htg10:
54: gb_htg11:
55: gb_htg12:
56: gb_htg13:
57: gb_htg14:
58: gb_in3:
59: gb_htg15:
60: gb_htg16:
61: gb_htg17:
62: em_htg4:
63: em_htg5:
64: em_htg6:
65: em_htg7:
66: em_hum6:
67: gb_htg18:
68: gb_htg19:
69: gb_htg20:
70: gb_htg21:
71: gb_htg22:
72: gb_htg23:
73: gb_htg24:
74: gb_htg25:
75: gb_htg26:
76: gb_htg27:
77: gb_htg28:
78: gb_htg29:
79: gb_htg30:
80: gb_htg31:
81: gb_vil:
82: gb_vil2:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	4106	98.6	4165	39	AF173155	AF173155 Homo sapi
2	3524	84.6	3989	50	AF157624	AF157624 Homo sapi
3	1107	26.6	178064	54	AC023311	AC023311 Homo sapi
4	306	7.3	77043	71	AC027757	AC027757 Homo sapi
c 5	306	7.3	178064	54	AC023311	AC023311 Homo sapi
6	268	6.4	184655	42	AC009581	AC009581 Homo sapi
c 7	125	3.0	77043	71	AC027757	AC027757 Homo sapi
8	38	0.9	4368	12	RNU76717	U76717 Rattus norv
c 9	22	0.5	65674	68	AC026980	AC026980 Homo sapi
10	22	0.5	146547	10	HS11B24	AL022394 Homo sapi
11	22	0.5	170891	41	AC002118	AC002118 Homo sapi
12	22	0.5	171741	40	AL138917	AL138917 Homo sapi
c 13	22	0.5	177163	10	HS134E15	AL022067 Human DNA
14	22	0.5	193756	56	AC024189	AC024189 Homo sapi
c 15	22	0.5	205746	32	AL133509	AL133509 Homo sapi
c 16	22	0.5	286758	39	AC006449	AC006449 Homo sapi
c 17	21	0.5	1798	8	AF074484	AF074484 Candida t
18	21	0.5	55077	42	AC013601	AC013601 Homo sapi
c 19	21	0.5	89712	9	AP001417	AP001417 Homo sapi
c 20	21	0.5	100000	9	AP000019	AP000019 Homo sapi
c 21	21	0.5	100000	9	AP000160	AP000160 Homo sapi
c 22	21	0.5	108049	79	AC016617	AC016617 Homo sapi
23	21	0.5	157061	73	AC011443	AC011443 Homo sapi
c 24	21	0.5	161801	54	AC008564	AC008564 Homo sapi

Query Match		84.6%;	Score 3524;	DB 50;	Length 3989;
Best Local Similarity		99.8%;	Pred. No. 0;		
Matches 3824;		Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;
Qy	246	ctgatagaaggagacttccatcaataacagcaacaaacctacttatcatatagagaatc	305		
Db	69	CTGATAGAAGGAGACTTTCATCATTAACAGACAAACCCACTTATCTATAGAGAGATC	128		
Qy	306	caagaaccacaagagtcaggttctctctctgctgaagaatacaacagaccttctcta	365		
Db	129	CAAGAACCACAGAGTCAGGTTCTTTCTCTGCTCGAAGNATCAACAGACCTTCTCTA	188		
Qy	366	gtaccacaaaagaacagcctcggaccgagagaaactgaagacagtggttaacaga	425		
Db	189	GCTACCAAAAAGAACAGCCTCGACCGCAGAGAAACTGAAGACACAGTGTGTTAACAGA	248		
Qy	426	aggcggagatctattctgttcttaataaggagtgaaatctgcccagatgaagtggtgca	485		
Db	249	AGCGGAGATCTATTCTGTTTCTTAATGGAGTGAATACTGCCAGATGAAGTGTGCA	308		
Qy	486	gaggtcgtggcaaatcatgtgaagtattttaagtcgagtggtgtcaggaagctgtctgg	545		
Db	309	GAGGCTGTGCAATCATGTGAAGTATTTAAAGTCGAGTGTGTGAGGAAGCTGTCTGG	368		
Qy	546	gaagccttcaggacttttgggacgcactcctggggtgaggaatatcatattactgatg	605		
Db	369	GAAGCCTTCAGGACTTTTGGGATCGACTTCTCGGGGTGAGGAATATCATTAATTGGGATG	428		
Qy	606	aatttgtgagatggagtcacaagtatatattgaatatgggcacaaatatttagtaatct	665		
Db	429	AATTTGTGTGAGATGGAGTFCACAAGTATATTTGAAATGGGCACAAAATTTTAGTGAATCT	488		
Qy	666	gtggaacatagaagcttaatacgaacaactgaacttatcaaggaactgtaagcagc	725		
Db	489	GTGGAACATAGAAGCTTAATCAAGAAGAACTGACTTATGCAAGGAAGAACTGTGAAGAGC	548		
Qy	726	tctgaactgtctctccagttcctgtgtggtgacttcaactcgaagtgagagacactactctc	785		
Db	549	TCGTGAATGCTTCTCCAGTTCCTGTGTGGTGATACTTCAACATTTGGGAGACACTACTCTC	608		
Qy	786	agtgttccacatccagaggtgagcgcctatgaaggtgcctcagagagcagcttggaagg	845		
Db	609	AGTGTTCACATCCAGAGGTGGAGCCCTATGAAGGTGCCTCAGAGAGCAGCTTGGAAGG	668		
Qy	846	ccagagagagattagcaatgaattggaatgtgtagaagaagccacaaaaccagca	905		
Db	669	CCAGAGGAGAGTATTAGCAATGAATTTGAGAAATGTGATGAAGAGCCACAAAACCCAGCA	728		
Qy	906	ggtgaacagattgcagaattcagatccaccttttgggagcagtcagcaggaagaacta	965		
Db	729	GGTGAACAGATTGCGAATTCAGTATCCACTTTTGGGGAGAGCAGTACAGGGAAGAACTA	788		
Qy	966	caggattctccagcttccaccaccagcaccttgaaagaataattatttcagaggttgaa	1025		
Db	789	CAGGATTCTCCAGCTTTCACCACCAGCACCTTGAAGAAATTTATTTCAGAGGTTGAA	848		
Qy	1026	aatgcattactgggtttaccagcgtacaagaaattcgtgtacttgtaatttagtcccc	1085		
Db	849	AATGCATTTACTGGGTTACCAGGCTACAGGTAACAGGAATTCGTGACTTGAATTTAGGTCGCC	908		
Qy	1086	aaggaaaatgacagtggtgctagattgtttactatgcagttaccttcaatggtgagggcatc	1145		
Db	909	AAGGAAAAATGACAGTGGCGTAGATGTTTACTATGTCAGTACCTTCAATGGTGAGGCCATC	968		
Qy	1146	agcaataccacctgggacctcattagccttcactcccaacaggtggaaaaccatggccctt	1205		
Db	969	AGCAATACCACTGGGACCTCATTAGGCTTCACTCCCAACAGGTGGAAAACCATGGGCTT	1028		
Qy	1206	gtggaaactggatgataaacccactgtgtttatacaatcatcagtaactcagagattattt	1265		
Db	1029	GTGGAATGGATGATAAACCCACTGTTGTTTATACAATCAGTAACTTCAGAGATTATATT	1088		
Qy	1266	gctgagacattgacgagaatatttttctgctggggaactcttctcttgaaatccagatccctgat	1325		

Db	1089	GCTGAGACATTGCGACGAAATTTTGTGGGAACTCTTCTGAAATCCAGATCCTGAT	1148		
Qy	1326	tectgcagcttatacaatgtgagagagtttgcgtcaccacaaactgaagatcagttgg	1385		
Db	1149	TCCCTGCGAGCTTATCAATGTGAGAGAGTTTGGCTCACCACAACTGAAGATCTAGTTGG	1208		
Qy	1386	aacacccaaagtccaagttcttcagggcaacgcgtcatctattcttggaataacaccttcaa	1445		
Db	1209	AACACCCAAAGTCAAGTCTTCAGGCAACGGCGTCATCTATTTCTGGATAATACCTTTCAA	1268		
Qy	1446	gctgatggccctcagcagatgaatccatcacccagtcagttattccaccactgatttcagc	1505		
Db	1269	GCTGATGGCCCTCAGCAGATGAATCCATACACAGCAGTATTTCCACCACCTGATTTCAGC	1328		
Qy	1506	tctggtcctcctcagccacitggcaggaactcctgttcagaaagtccttgggtgatitca	1565		
Db	1329	TCTGTCTCTCCTCAGCCACTGGCAGGAACTCTGGTCAGAAAGTCTTTGGGTGATTTA	1388		
Qy	1566	gtgtcacacacaaattagcctttccctcgaagatgggcctcagctcttccccagaggtt	1625		
Db	1389	GTGTCTACACACAAATTAGCCTTTCCCTCGAAGATGGGCTTCAGCTCTCCCCACAGGTT	1448		
Qy	1626	ttagaggttagcagcttgacttcttcttctcaccocggcagtgcttccagactggttg	1685		
Db	1449	TTAGAGGTTAGCAGCTTGACTCTTCTATCTGTACCCCGGCAGTGTCTCAGACTGGCTTG	1508		
Qy	1686	cctgtggtcctgaggaaggaacttctggtatctcacttgtagaagatggattagccaat	1745		
Db	1509	CCTGTGGCTCTGTAGAAAGGACTTCTGGATCTCACTTGGTAGAAGATGGATTAGCCAA	1568		
Qy	1746	gtgaagagtcagagaatttcttcttcttcttcttcttcttcttcttcttcttcttcttca	1805		
Db	1569	GTGGAAGAGTCAGAGATTTTCTTCTATTTGATTCATTCCTTCAAGTTCATTCTCAA	1628		
Qy	1806	cctgtgcaaaagaacaaatccatccatggaagactctggtcttcttcttcttcttcttctca	1865		
Db	1629	CCTGTGCCAAAAGAAACAAATACCATCCATGGAAGACTCTGTGTCTCTTAACATCTTCA	1688		
Qy	1866	ccatatctgacctcttctataccttttggcttggactccttgacctcaaaagtcacaaagac	1925		
Db	1689	CCATATCTGACCTCTCTTATACCTTTTGGCTGGGACTCCTTGACCTCCAAAGTCAAGAC	1748		
Qy	1926	caatataaagtgcagcccttctcctgcagatgcacatccatggaaagagtaataattgac	1985		
Db	1749	CAATTAAGATGAGCCCTTCTGCGCAGATGCATCCATGGAAAAAGATTAATATTGAC	1808		
Qy	1986	ggtggttttaggttcagggctctgggcaaaaggtagatctgattacttggccaaggagtgag	2045		
Db	1809	GGTGGTTTAGGTTACAGGCTCTGGGCAAAAGGTAGATCTGATTCTGGCCATGGAGTGAG	1868		
Qy	2046	acttcacagagaagagcgcggaacacactgtccaaagcgcgtggcttgaagatgatattca	2105		
Db	1869	ACTTCATCAGAGAAGAGCGCTGAACCACTCTCAAGCGGTGGCTTGAAGATGATGATCA	1928		
Qy	2106	cttttgcagctgagattgaagacaaactagtttttagtttgacacaaatggattccaca	2165		
Db	1929	CTTTTGGCAGCTGAGATTGAACAGAAAGAACTAGTTTGTAGTTCACAAATGGATTTCCA	1988		
Qy	2166	gacaaaattagtaagcactcaaaatatagaatgatgacagatccacacacttccacag	2225		
Db	1989	GACCAATTTAGTAGCACTCAAAATATGTACATGTACAGATCCATACACTTCCAGAG	2048		
Qy	2226	gaagagcctcttagtgggctgctgtgcccactcttcgagatactgcagctgaactcgcg	2285		
Db	2049	GAAGAGCCTCTTAGTGGGCTCTGTGCCCCATCTTCGACAGATCTGCAGCTGAATCTGG	2108		
Qy	2286	tctctaacctcccccagacatatcagaagtacccgtgtgttgatgtctcagttacc	2345		
Db	2109	TCTCTAACCTCCCCAGCACATATCAGAAGTACCTGGTGTGTGATGATTACTCAGTTACC	2168		
Qy	2346	aagcacctcttatactgaca tctgttagcaatctctgctctactgataaatcagatcag	2405		

Ds	2169	AAAGCACCTCTTATACATGACATCTGTAGCAATCTCTGCTCTACTGATATAATACAGATCAG	2228
QY	2406	gcagatgcacccctaaaggagagatgaacaaaattactgactcatcaacatataatgg	2465
Ds	2229	GCAGATGCCATCCTTAAGGGAGGATATGGAACAAATTAAGTGCATCCAACTATGAATGG	2288
QY	2466	tttgacagtgaggtttcaatggtaaaagccagatagcaaaactttgggactatatggca	2525
Ds	2289	TTTGACAGTGAAGTTCAATGGTAAGCCAGATATGCAAACTTTGGGACTATATTGCCA	2348
QY	2526	gaatcagagagagtttgacaaagaacttctccctagagaaaattgtccagagacatatgg	2585
Ds	2349	GAATCAGAGAGAGTTTGCAAGAACTTCTCCCTAGAGAAAATTGTCAGAGACATATTG	2408
QY	2586	gcaagtacacacagagtgctgaaggtctggttatctgtatctgtgaacagctctaccaaatgg	2645
Ds	2409	GCAAGTACACCACAGAGTGTCTGACAGGTCTGTGTATCTGTGACACAGTCTACCAAAATGG	2468
QY	2646	cctccaaaccacaattccaccctgctagaggatgaagtaattatgggtgtacagatatatt	2705
Ds	2469	CCTCCAAACCACAATCTCCACCCTGTAGAGGATGAAGTAATTAATGGGTGACAGGATATT	2528
QY	2706	tcgtagaacgagccggatagggacagattactatcagcctgagcaagtcccaagagcaa	2765
Ds	2529	TCGTTAGAACTGGACCGGATAGGCACAGATTACTATCAGCCTGAGCAAGTCCCAAGAGCAA	2588
QY	2766	aatggcaaggttgtagttatgtgaaatgtcaacaagtgttcaactccacagagatggtt	2825
Ds	2589	AATGGCAAGGTGGTGTATGTGGAATGTCAACAAGTGTTCACCTCCACAGAGATGGTT	2648
QY	2826	agtggtggttggccacagaaggagagatgacttgagtgtataccagacttcaggagct	2885
Ds	2649	ACTGTGGCTTGGCCCAAGAGGAGAGATGACTTGAGTTATACCAGACTTCAGGAGCT	2708
QY	2886	ttggtggttttcttcagcctccgagtgactaacaatgatgttttcagaagatgtgtttaa	2945
Ds	2709	TTGGTGGTGTCTTCAGCCTCCGAGTGAATAACATGATGTTTTTCAGAAGATCTGTTTAAT	2768
QY	2946	aaaaactccttgagatataaagccctgagcaaaagattcttgaattgctggttccctat	3005
Ds	2769	AAAAACTCCTTTGGAGATATAAGCCCTGGAGCAAGATTTCTTGAATTTGCTGTTCCCTAT	2828
QY	3006	ctccagtaaatctcacggggttccagaaacttagaaactcctcaacttcagaaatggcagc	3065
Ds	2829	CTCCAGTCAAAATCTCACGGGGTTCCAGAACTTAGAAATCCTCACTTCAGAAATGGCAGC	2888
QY	3066	attggtggaacagtcgaatgaagtttgccaattctgtccctcctaagctcaacaatgoc	3125
Ds	2889	ATTGTGGTGAACAGTCGAATGAAGTTTGCAATTTCTGTCCTCTCAACCTCAACATGCG	2948
QY	3126	gtgtacatgattctggaagactttgtaccactgectacaataccaatgaacttggctatt	3185
Ds	2949	GTGTACATGATTCTGGAAGACTTTTGTACCACCTGGCTGCTACAAATACCATGAACCTTGGCTATT	3008
QY	3186	gataaataactcttgatgtggaacaggatgagccaacccctgagtggaagcaagtcagatgc	3245
Ds	3009	GATAAATACTCTCTGTGTGGAAATCAGGTGATGAAGCCCAACCCCTTGCAGTTTCAGGCC	3068
QY	3246	tgtaatgaatttccagagtgctgtgtcaccctcgagtggaagcaacccctgcaagttcaggcc	3305
Ds	3069	TGTAATGAATTTTCAGATGCTGTGTCAACCCCTGGAGTGGAGAACCAAGTGCAGATGC	3128
QY	3306	ttccctggataccttgatgtggaagaagccctgtcagagttctctgtgacctacagcct	3365
Ds	3129	TTCCCTGGATACCTCAGTGTGGAAGAACGGCCCTGTACAGAGTCTCTGTGACCTACAGCCT	3188
QY	3366	gacttctgtgaatgataggaaagtgtgaactattgctcctggcgacagggccatttgaag	3425
Ds	3189	GACTCTCTGTTGAATGATGGAAGTGTGACATTATGCTTGGCAGCGGGCCATTGTAGG	3248
QY	3426	tgccgggtgggtgagaactgggtgaccgagcgaagcactgtgagaaattgtgtctgag	3485
Ds	3249	TGCCGGGTGGGTGAGAACTGGTGGTACCAGGACCAAGCACTGTGAGGAATTTGTGCTGAG	3308
QY	3486	cccgatcatcattagggcatcactattgctccgtggttggtgacttcttcttcttctgct	3545
Ds	3309	CCCGTGATCATAGGCATCACTATTGCTCCGTGGTGGACTTCTTCTCATCTTTCTTGCT	3368
QY	3546	atcatctacttcttcatcaggactcttcaagcacacacatgacagggagtgaagagagagt	3605
Ds	3369	ATCATCTACTTCTTTCATCAGGACTCTTCAAGCACACCATGACAGGAGTGAAGAGAGAGT	3428
QY	3606	cccttcagtggtccacagagcctgacagcctctcatctatttattgagaatgctgtgaag	3665
Ds	3429	CCCTTCAGTGGTCCAGCAGCGCCTGACGCTCTCATCTATTGAGAATGCTGTGAAG	3488
QY	3666	tacaaccccggtgatgaaggtcacagggctgagatggaagatgagagaccctatcct	3725
Ds	3489	TACAACCCCGTGTATGAAAGTCAAGGGCTGGATGTGAGAAGTATGAGGGACCCCTATCCT	3548
QY	3726	cagcatcctcttctacagctctgtagcggagacgtgattggtgggctgagcagagaagaa	3785
Ds	3549	CAGCATCCCTTCTACAGCTCTGCTAGCGGAGAGCTGATTGGTGGCTGAGCAGAGAAGAA	3608
QY	3786	atcagcacatgatatgagcagtgagcgtttccagagagagaaattcaagagagaatgaga	3845
Ds	3609	ATCAGACAGATGTATGAGAGCAGTGAGCTTCCAGAGAGGAAATTCAGAGAGAATGAGA	3668
QY	3846	gttttgaaactgtatcccaatgactcctgagtttgagcttttggagagcaacaagt	3905
Ds	3669	GTTTTGAAGTGTATGCCAATGATCCTGAGTTGCGAGCTTTTGTGAGAGAACCAAGTG	3728
QY	3906	gaagaggtttaacccaaactcctgttctgaaactgattagaagcctgggagaagatggaga	3965
Ds	3729	GAAGAGTTTAAACCAAACTCCTGTTCTGAACTGATTAGAACCTTGAGAACATGGAGA	3788
QY	3966	ttactgttacttatctatataataactgaactgagattttaaactggttggagaagagnt	4025
Ds	3789	TTACTTGTACTTATGTCATATAAATTAAGCTGATTTTAAACACTGTGTGAAGAAGAGTT	3848
QY	4026	tictatgaaaaaattaaatatagggcacactgttttttttctcagcttaag	4075
Ds	3849	TTCTATGAAAAAATTAATATAGGGCACACTGTGTTTTTTTTCAGCTTAAG	3898
RESULT	3		
AC023311			
LOCUS	178064 bp	DNA	HTG
DEFINITION	Homo sapiens clone RP11-571M4, WORKING DRAFT SEQUENCE, 20 unordered pieces.		03-MAR-2000
ACCESSION	AC023311		
VERSION	AC023311.3	GI:7144985	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 178064)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Camporeale,A., Castle,A., Choepey,Y., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M., Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEvan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,		

1986	Db	TCATTCACTCAACCTGTGCCAAAAGAAACAATACCATCATCGAAGACTCTGATGTGTC	2045
1987	Qy	ttaacatcttcaccatactgacctctctctataccttttggcttggaacctcttgacctcc	1913
1988	Db	TTAACATCTTCACCATATCTGACCTCTTCTATACCTTTTGGCTTGACCTCCTTGACCTCC	2105
1989	Qy	aaagtcaaaagaccaattaaaaagtgagcccttctctgcccagatgcatccatggaataagag	1973
1990	Db	AAAGTCAAGACCAATTAAAGTGAGCCCTTCTCGCAGATGCATCCATGGAAAAAGAG	2165
1991	Qy	ttaattattgacgggtgggttaggtttcagggtctgggcaaaagtagatctgattacttgg	2033
1992	Db	TTAATATTGACGGTGGTTAGTTCAGGGTCTGGCAAAAGGTAGATCTGTATTACTTGG	2225
1993	Qy	ccatggagtgaacttcacagaagaagcgcgcgaaccactgtccaaagccgtggctgaa	2093
1994	Db	CCATGGAGTGAGACTTCATCAGAGAAGACGGTGAACCACTGTCCAAAGCGGTGGCTTGA	2285
1995	Qy	gatgatattcaacttttggccagctgagattgaagacaagaactagtcttttagttgacaaa	2153
1996	Db	GATGATGATTCACCTTTGGCAGCTGAGATTGAAGACAAGAAACTAGTTTTAGTTGACAAA	2345
1997	Qy	atggattccacagacacaaattagtaagaactcaaaatagaacatgatgacagatccaca	2213
1998	Db	ATGGATTCCACAGACCAAAATTAGTAGCACTCAAAATATGAACATGATGACAGATCCATA	2405
1999	Qy	cactttccagaggaaagacctcttagtggccctgctgtgcccatcttcgcagatactgca	2273
2000	Db	CACITTCAGAGGAGAGCCTCTTAGTGGCCCTGCTGTGCCCATCTTCGCAGATATCGCA	2465
2001	Qy	gctgaatcgcgtctctaacctccccacacatcatcagaagtacctgggttgatgat	2333
2002	Db	GCTGAATCTGGCTCTTAACCTCCCAAGCACATATCAGAACTACCTGGTGTGTGATGAT	2525
2003	Qy	tgctcagttaccaaaagacctcttatactgacatctgtagcaatctctgcctctactgat	2393
2004	Db	TACTCAGTTACCAAGACACCTCTTATACGTACATCTGTAGCAATCTCGCCTCTACTGAT	2585
2005	Qy	aaatcagatcaggcgagatgccatcctaaaggaggatattgaaacaaatattcagagtcaccc	2453
2006	Db	AAATCAGATCAGGCAGATGCCATCTTAAGGGAGGATATGGACAAATTACTGAGTCATCC	2645
2007	Qy	aactatgaattgggtttgacagtgaggtttcaatggtaaaagccagatatgcaaacctttgtgg	2513
2008	Db	AACTATGAATGGTTTGACAGTGAGGTTTCAATGGTAAAGCCAGATATGCAAACTTTGTGG	2705
2009	Qy	actatatgccagaatcagagagaggtttgacaagaactcttccctagagaataatgtcc	2573
2010	Db	ACTATATTGCCAGAAATCAGAGAGAGTTTGGCAAGAACTTCTTCCCTAGAGAAATTTGTC	2765
2011	Qy	agagacatattggcaagtacacacagagtgctgcagagctctgggttattctgtgacacag	2633
2012	Db	AGAGACATATTGGCAAGTACACACAGAGTGTGACAGGCTCTGGTTATCTGTGACACAG	2825
2013	Qy	tctaccaaatgtcctcccaaccacaatctccacctgctagaggatgaagtaattatgggt	2693
2014	Db	TCTACCAAAATGGCTTCCACCAACCAATCTCCACCTGCTAGAGGATGAAGTAATATGGGT	2885
2015	Qy	gtacaggatatctcgttagaactggaaccggaaggagagattactatcagccctggagcaa	2753
2016	Db	GTACAGGATATTTCCGTTAGAACTGGACCGGATAGGCACAGATTACTATCAGCCTGAGCAA	2945
2017	Qy	gtccaagagcaaaatggcaagtttggttagttatgtggaatgtccaacagtttcaactcc	2813
2018	Db	GTCCAAGAGCAAAAATGGCAAGGTTTGATTTATGTGGAAATGTCAACAAGTTGTCTACTCC	3005
2019	Qy	acagagatgggttagtggcttggccccaagaggagagatgacttgattatacccca	2873
2020	Db	ACAGAGATGGTTAGTGTGGCTTGGCCCAACAGAGGAGGAGATGACTTGATTATACCCAG	3065
2021	Qy	acttcaggagctttgggtgggtttcttcagccctccgagtgactaaacatgatgttttcagaa	2933
2022	Db	ACTTCAGAGAGCTTTGGTGGTTTCTTCAGCCTCCGAGTGACTAACTATGATGTTTCAGAA	3125


```
* 1614 1713: gap of 100 bp
* 1714 2482: contig of 769 bp in length
* 2483 2582: gap of 100 bp
* 2583 3347: contig of 765 bp in length
* 3348 3447: gap of 100 bp
* 3448 4243: contig of 796 bp in length
* 4244 4343: gap of 100 bp
* 4344 5088: contig of 745 bp in length
* 5089 5188: gap of 100 bp
* 5189 5968: contig of 780 bp in length
* 5969 6068: gap of 100 bp
* 6069 6832: contig of 764 bp in length
* 6833 6932: gap of 100 bp
* 6933 7682: contig of 750 bp in length
* 7683 7782: gap of 100 bp
* 7783 8560: contig of 778 bp in length
* 8561 8660: gap of 100 bp
* 8661 9445: contig of 785 bp in length
* 9446 9545: gap of 100 bp
* 9546 10318: contig of 773 bp in length
* 10319 10418: gap of 100 bp
* 10419 11205: contig of 787 bp in length
* 11206 11305: gap of 100 bp
* 11306 12077: contig of 772 bp in length
* 12078 12177: gap of 100 bp
* 12178 12959: contig of 782 bp in length
* 12960 13059: gap of 100 bp
* 13060 13855: contig of 796 bp in length
* 13856 13955: gap of 100 bp
* 13956 14719: contig of 764 bp in length
* 14720 14819: gap of 100 bp
* 14820 15602: contig of 783 bp in length
* 15603 15702: gap of 100 bp
* 15703 16478: contig of 776 bp in length
* 16479 16578: gap of 100 bp
* 16579 17357: contig of 779 bp in length
* 17358 17457: gap of 100 bp
* 17458 18237: contig of 780 bp in length
* 18238 18337: gap of 100 bp
* 18338 19127: contig of 790 bp in length
* 19128 19227: gap of 100 bp
* 19228 20029: contig of 802 bp in length
* 20030 20129: gap of 100 bp
* 20130 20899: contig of 770 bp in length
* 20900 20999: gap of 100 bp
* 21000 21786: contig of 787 bp in length
* 21787 21886: gap of 100 bp
* 21887 22657: contig of 771 bp in length
* 22658 22757: gap of 100 bp
* 22758 23535: contig of 778 bp in length
* 23536 23635: gap of 100 bp
* 23636 24435: contig of 800 bp in length
* 24436 24535: gap of 100 bp
* 24536 25330: contig of 795 bp in length
* 25331 25430: gap of 100 bp
* 25431 26223: contig of 793 bp in length
* 26224 26323: gap of 100 bp
* 26324 27092: contig of 769 bp in length
* 27093 27192: gap of 100 bp
* 27193 27957: contig of 765 bp in length
* 27958 28057: gap of 100 bp
* 28058 28832: contig of 775 bp in length
* 28833 28932: gap of 100 bp
* 28933 29687: contig of 755 bp in length
* 29688 29787: gap of 100 bp
* 29788 30558: contig of 771 bp in length
* 30559 30658: gap of 100 bp
* 30659 31425: contig of 767 bp in length
* 31426 31525: gap of 100 bp
* 31526 32305: contig of 780 bp in length
* 32306 32405: gap of 100 bp
* 32406 33177: contig of 772 bp in length
* 33178 33277: gap of 100 bp
```

```
* 33278 34048: contig of 771 bp in length
* 34049 34148: gap of 100 bp
* 34149 34934: contig of 786 bp in length
* 34935 35034: gap of 100 bp
* 35035 35793: contig of 759 bp in length
* 35794 35893: gap of 100 bp
* 35894 36658: contig of 765 bp in length
* 36659 36758: gap of 100 bp
* 36759 37545: contig of 787 bp in length
* 37546 37645: gap of 100 bp
* 37646 38403: contig of 758 bp in length
* 38404 38503: gap of 100 bp
* 38504 39296: contig of 793 bp in length
* 39297 39396: gap of 100 bp
* 39397 40179: contig of 783 bp in length
* 40180 40279: gap of 100 bp
* 40280 41037: contig of 778 bp in length
* 41038 41137: gap of 100 bp
* 41138 41919: contig of 762 bp in length
* 41920 42019: gap of 100 bp
* 42020 42800: contig of 781 bp in length
* 42801 42900: gap of 100 bp
* 42901 43677: contig of 777 bp in length
* 43678 43777: gap of 100 bp
* 43778 44549: contig of 772 bp in length
* 44550 44649: gap of 100 bp
* 44650 45427: contig of 778 bp in length
* 45428 45527: gap of 100 bp
* 45528 46321: contig of 794 bp in length
* 46322 46421: gap of 100 bp
* 46422 47201: contig of 780 bp in length
* 47202 47301: gap of 100 bp
* 47302 48095: contig of 794 bp in length
* 48096 48195: gap of 100 bp
* 48196 48979: contig of 784 bp in length
* 48980 49079: gap of 100 bp
* 49080 49826: contig of 747 bp in length
* 49827 49926: gap of 100 bp
* 49927 50778: contig of 852 bp in length
* 50779 50878: gap of 100 bp
* 50879 51649: contig of 771 bp in length
* 51650 51749: gap of 100 bp
* 51750 52522: contig of 773 bp in length
* 52523 52622: gap of 100 bp
* 52623 53379: contig of 757 bp in length
* 53380 53479: gap of 100 bp
* 53480 54258: contig of 779 bp in length
* 54259 54358: gap of 100 bp
* 54359 55101: contig of 743 bp in length
* 55102 55201: gap of 100 bp
* 55202 55979: contig of 778 bp in length
* 55980 56079: gap of 100 bp
* 56080 56873: contig of 794 bp in length
* 56874 56973: gap of 100 bp
* 56974 57726: contig of 753 bp in length
* 57727 57826: gap of 100 bp
* 57827 58634: contig of 808 bp in length
* 58635 58734: gap of 100 bp
* 58735 59545: contig of 811 bp in length
* 59546 59645: gap of 100 bp
* 59646 60413: contig of 768 bp in length
* 60414 60513: gap of 100 bp
```

Query Match 7.3%; Score 306; DB 71; Length 77043;
Best Local Similarity 100.0%; Pred. No. 7.4e-166;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1430 ggataataccttttaagctgcatggccctcagcagagaatccatccaccagcagattccc 1489
|||||
Db 24639 GGATAATACCTTTTCAAGCTGCATGGCCCTCAGCAGATGAATCCATCCACGAGATATCC 24698
|||||
QY 1490 accacttgattcagctctggtctcctcagccactggcagggaactcgtggtcagaag 1549
|||||


```

misc_feature 10096..14245
/note="assembly_fragment"
misc_feature 14346..17761
/note="assembly_fragment"
clone_end:SP6
vector_side:left
17862..24067
/note="assembly_fragment"
24168..29654
/note="assembly_fragment"
29755..35767
/note="assembly_fragment"
35868..42494
/note="assembly_fragment"
42595..49216
/note="assembly_fragment"
clone_end:T7
vector_side:left
49317..59837
/note="assembly_fragment"
59938..68982
/note="assembly_fragment"
69083..77488
/note="assembly_fragment"
77589..88040
/note="assembly_fragment"
88141..100197
/note="assembly_fragment"
100298..114772
/note="assembly_fragment"
114873..132014
/note="assembly_fragment"
132115..156372
/note="assembly_fragment"
156473..178064
/note="assembly_fragment"
BASE COUNT 56338 a 33224 c 33032 g 53564 t 1906 others
ORIGIN
Query Match 7.3%; Score 306; DB 54; Length 178064;
Best Local Similarity 100.0%; Pred. No. 7.7e-166;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1430 gataataactttcaagctgctggtccctcagcagatgaatccatccaccagcagattcc 1489
|||||
Db 36685 GGATAAATCTTTCAAGCTGCATGGCCCTCAGCAGATGAATCCATCCACGAGATATCC 36626
|||||
QY 1490 accacttgattcagctggtctccctcagcactggcagggaactctggtcagaaag 1549
|||||
Db 36625 ACCACTTGATTTCAGCTCTGGTCTCCCTCAGGCACATGGCAGGGAATCTGGTCAGAAAG 36566
|||||
QY 1550 tctcttgggtgattagtgctctacacaaattagccttccctcgaagatgggcctcag 1609
|||||
Db 36565 TCCTTGGGTGATTAGTGTCTACACAAATTAGCTTCCCTCGAAGATGGGCTTCAG 36506
|||||
QY 1610 ctcttccagaggttttagaggttagcagcttgactcttctcattctcaccctcaggt 1669
|||||
Db 36505 CTTCTCCAGAGGTTTGAAGATTAGCAGCTTGACTCTTCATCTCTCACCCTGAGGT 36446
|||||
QY 1670 gcttgaactggttccctggtctctgaggaagactcttgatctcacttggtaga 1729
|||||
Db 36445 GCTTCAGACTGGCTTCCCTGTGGCTCTGAGGAAGAGGATCTTGGATCTCATTGGTAGA 36386
|||||
QY 1730 agatgg 1735
|||||
Db 36385 AGATGG 36380
|||||
RESULT 6
AC009581 184655 bp DNA HTG 13-DEC-1999
LOCUS Homo sapiens chromosome 4 clone RP11-144C15 map 4, WORKING DRAFT
DEFINITION

```

ACCESSION
AC009581.3 GI:6563621
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 184655)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens chromosome 4, clone RP11-144C15

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 184655)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Miloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 13, 1999 this sequence version replaced gi:6437680.

COMMENT

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1725

Center clone name: 144_C_15

----- Summary Statistics

Sequencing vector: M13: M77815; 100% of reads

Chemistry: Dye-primer-amersham; 4% of reads

Assembly: Dye-terminator Big Dye; 96% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 156115 bases at least Q40

Consensus quality: 173950 bases at least Q30

Consensus quality: 181135 bases at least Q20

Insert size: 188000; agarose-fp

Insert size: 184655; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 4.3 in Q20 bas.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 12 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 433: contig of 433 bp in length

* gap of unknown length

* 434 3113: contig of 2680 bp in length

* gap of unknown length

* 3114 6298: contig of 3185 bp in length

* gap of unknown length

* 6299 10330: contig of 4032 bp in length

* gap of unknown length

* 10331 15690: contig of 5360 bp in length

* gap of unknown length

* *

```
* 15691 20515: contig of 4825 bp in length
* gap of unknown length
* 20516 34840: contig of 14325 bp in length
* gap of unknown length
* 34841 54827: contig of 19987 bp in length
* gap of unknown length
* 54828 74530: contig of 19703 bp in length
* gap of unknown length
* 74531 93245: contig of 18715 bp in length
* gap of unknown length
* 93246 116394: contig of 23149 bp in length
* gap of unknown length
* 116395 184655: contig of 68261 bp in length.
FEATURES
  Location/Qualifiers
    1..184655
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="4"
      /map="4"
      /clone="RP11-144C15"
      /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 64570 a 35140 c 33570 g 51373 t 2 others
ORIGIN

Query Match 6.4% Score 268 DB 42 Length 184655;
Best Local Similarity 100.08; Pred. No. 1e-143;
Matches 268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 ttgaagaagacaaccattttttccgcctaattataatggttttgaagtgtgttcca 68
Db 112476 TTTGAAGGACAACCATTTTCTTCCGCTAATTTATATGTTTGAAGTGTGTCTCA 112535

Qy 69 tctcaaacatagacttttaaatgttaggtcttctctataactcttctgttatgaagt 128
Db 112536 TTCTCAACATAGACTTTTAAATGTTAGGTCTTCTCTATAACTCTTGTATTGGAAGTT 112595

Qy 129 tcaggatttgacactcaataaagattctctctctctctctctctctctctctctcttggcc 188
Db 112596 TCAAGGATTGACACTCAATTAAGATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTGGCC 112655

Qy 189 caaatgattgttctctctctttttgggaagattctctctgggtatttggatattgtcctg 248
Db 112656 CAATGATTGTTTCTCTCTTTTGGGAGATTCTCTGGGTATTGTGATATTGTCTCTG 112715

Qy 249 atagaagagagacttccatcaataacag 276
Db 112716 ATAGAAGGAGACTTCCATCATTAACAG 112743

RESULT 7
AC027757/c 77043 bp DNA HTG 01-APR-2000
LOCUS Homo sapiens chromosome 3 clone RP11-663C11 map 3, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC027757
VERSION AC027757.1 GI:7382583
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 77043)
  Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
  Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
  Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
  Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
  Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
  Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
  Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
  Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
  Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
  Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
  Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
  McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
  Meidrum,J., Meneus,L., Mirava,T., Miranda,C., Mlenga,V., Morrow,J.,
  Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
  O'Neil,D., Olivat,T.M., Oliver,J., Peterson,K., Pierre,N.,
  Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
  Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
  Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
  Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
  Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
  Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8272
Center clone name: 663_C_11
-----
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 763: contig of 763 bp in length
* 764 863: gap of 100 bp
* 864 1613: contig of 750 bp in length
* 1614 1713: gap of 100 bp
* 1714 2482: contig of 769 bp in length
* 2483 2582: gap of 100 bp
* 2583 3347: contig of 765 bp in length
* 3348 3447: gap of 100 bp
* 3448 4243: contig of 796 bp in length
* 4244 4343: gap of 100 bp
* 4344 5088: contig of 745 bp in length
* 5089 5188: gap of 100 bp
* 5189 5968: contig of 780 bp in length
* 5969 6068: gap of 100 bp
* 6069 6832: contig of 764 bp in length
* 6833 6932: gap of 100 bp
* 6933 7682: contig of 750 bp in length
* 7683 7782: gap of 100 bp
* 7783 8560: contig of 778 bp in length
* 8561 8660: gap of 100 bp
* 8661 9445: contig of 785 bp in length
* 9446 9545: gap of 100 bp
* 9546 10318: contig of 773 bp in length
* 10319 10418: gap of 100 bp
* 10419 11205: contig of 787 bp in length
* 11206 11305: gap of 100 bp
* 11306 12077: contig of 772 bp in length
* 12078 12177: gap of 100 bp
* 12178 12959: contig of 782 bp in length
* 12960 13059: gap of 100 bp
* 13060 13855: contig of 796 bp in length
* 13856 13955: gap of 100 bp
```

* 13956 14719: contig of 764 bp in length
* 14720 14819: gap of 100 bp
* 14820 15602: contig of 783 bp in length
* 15603 15702: gap of 100 bp
* 15703 16478: contig of 776 bp in length
* 16479 16578: gap of 100 bp
* 16579 17357: contig of 779 bp in length
* 17358 17457: gap of 100 bp
* 17458 18237: contig of 780 bp in length
* 18238 18337: gap of 100 bp
* 18338 19127: contig of 790 bp in length
* 19128 19227: gap of 100 bp
* 19228 20029: contig of 802 bp in length
* 20030 20128: gap of 100 bp
* 20130 20899: contig of 770 bp in length
* 20900 20999: gap of 100 bp
* 21000 21786: contig of 787 bp in length
* 21787 21886: gap of 100 bp
* 21887 22657: contig of 771 bp in length
* 22658 22757: gap of 100 bp
* 22758 23535: contig of 778 bp in length
* 23536 23635: gap of 100 bp
* 23636 24435: contig of 800 bp in length
* 24436 24535: gap of 100 bp
* 24536 25330: contig of 795 bp in length
* 25331 25430: gap of 100 bp
* 25431 26223: contig of 793 bp in length
* 26224 26323: gap of 100 bp
* 26324 27092: contig of 769 bp in length
* 27093 27192: gap of 100 bp
* 27193 27957: contig of 765 bp in length
* 27958 28057: gap of 100 bp
* 28058 28832: contig of 775 bp in length
* 28833 28932: gap of 100 bp
* 28933 29687: contig of 755 bp in length
* 29688 29787: gap of 100 bp
* 29788 30558: contig of 771 bp in length
* 30559 30658: gap of 100 bp
* 30659 31425: contig of 767 bp in length
* 31426 31528: gap of 100 bp
* 31528 32305: contig of 780 bp in length
* 32306 32405: gap of 100 bp
* 32406 33177: contig of 772 bp in length
* 33178 33277: gap of 100 bp
* 33278 34048: contig of 771 bp in length
* 34049 34148: gap of 100 bp
* 34149 34934: contig of 786 bp in length
* 34935 35034: gap of 100 bp
* 35035 35793: contig of 759 bp in length
* 35794 35893: gap of 100 bp
* 35894 36658: contig of 765 bp in length
* 36659 36758: gap of 100 bp
* 36759 37545: contig of 787 bp in length
* 37546 37645: gap of 100 bp
* 37646 38403: contig of 758 bp in length
* 38404 38503: gap of 100 bp
* 38504 39296: contig of 793 bp in length
* 39297 39396: gap of 100 bp
* 39397 40179: contig of 783 bp in length
* 40180 40279: gap of 100 bp
* 40280 41057: contig of 778 bp in length
* 41058 41157: gap of 100 bp
* 41158 41919: contig of 762 bp in length
* 41920 42019: gap of 100 bp
* 42020 42800: contig of 781 bp in length
* 42801 42900: gap of 100 bp
* 42901 43677: contig of 777 bp in length
* 43678 43777: gap of 100 bp
* 43778 44549: contig of 772 bp in length
* 44550 44649: gap of 100 bp
* 44650 45427: contig of 778 bp in length
* 45428 45527: gap of 100 bp
* 45528 46321: contig of 794 bp in length

* 46322 46421: gap of 100 bp
* 46422 47201: contig of 780 bp in length
* 47202 47301: gap of 100 bp
* 47302 48095: contig of 794 bp in length
* 48096 48195: gap of 100 bp
* 48196 48979: contig of 784 bp in length
* 48980 49079: gap of 100 bp
* 49080 49826: contig of 747 bp in length
* 49827 49926: gap of 100 bp
* 49927 50778: contig of 852 bp in length
* 50779 50878: gap of 100 bp
* 50879 51649: contig of 771 bp in length
* 51650 51749: gap of 100 bp
* 51750 52522: contig of 773 bp in length
* 52523 52622: gap of 100 bp
* 52623 53379: contig of 757 bp in length
* 53380 53479: gap of 100 bp
* 53480 54258: contig of 779 bp in length
* 54259 54358: gap of 100 bp
* 54359 55101: contig of 743 bp in length
* 55102 55201: gap of 100 bp
* 55202 55979: contig of 778 bp in length
* 55980 56079: gap of 100 bp
* 56080 56873: contig of 794 bp in length
* 56874 56973: gap of 100 bp
* 56974 57726: contig of 753 bp in length
* 57727 57826: gap of 100 bp
* 57827 58634: contig of 808 bp in length
* 58635 58734: gap of 100 bp
* 58735 59545: contig of 811 bp in length
* 59546 59645: gap of 100 bp
* 59646 60413: contig of 768 bp in length
* 60414 60513: gap of 100 bp

Query Match 3.0%; Score 125; DB 71; Length 77043;
Best Local Similarity 100.0%; Pred. No. 1.9e-60;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1430 ggataacatttcagctgcatggccctcagcagatgaatccatccacagcagattcc 1489
|||||
Db 22966 GGATATACCTTTCAGCTGCTGCTCCCTCAGCCACTGGCAGGAACTCTGGTCAGAAAG 22907
|||||

QY 1490 accactgtatttcagctgctgctcctcctcagccactgagcaggaactctgtgcagaag 1549
|||||
Db 22906 ACCACTTGATTTTCAGCTGCTGCTCCCTCAGCCACTGGCAGGAACTCTGGTCAGAAAG 22847
|||||

QY 1550 tccct 1554
|||||

Db 22846 TCCTT 22842

RESULT 8

RN076717
LOCUS RN076717 4368 bp mRNA ROD
DEFINITION Rattus norvegicus PG10.2 mRNA, complete cds.
ACCESSION U76717
VERSION U76717.1 GI:1667596
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 4368)
Wang, X., Brownstein, M.J. and Young, W.S. 3rd.
Sequence analysis of PG10.2, a gene expressed in the pineal gland
and the outer nuclear layer of the retina
Brain Res. Mol. Brain Res. 41 (1-2), 269-278 (1996)
JOURNAL 97038366
MEDLINE
REFERENCE 2 (bases 1 to 4368)
Wang, X., Brownstein, M.J. and Young, W.S.
Direct Submission
TITLE Submitted (30-OCT-1996) LCMR, NIMH, 36 Convent Drive, MSC 4068,
Bethesda, MD 20892-4068, USA



FEATURES

Location/Qualifiers
1. 4368
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
190. 3909
/gene="PG10.2"
190. 3909
/gene="PG10.2"
/note="expressed in the pineal gland and retina"
/codon_start=1
/protein_id="AAC52891.1"
/db_xref="GI:1667597"

/translation="MPLPLGRISLIGLILFTGGLNLSVSEETQDRMAHVAVLSPRES
TDLSPTREROLLDAETGRWRPRLRRRSILFPGVRCIPSDTFAEAVANHVYFAR
VQCEALWEAFRFDLPGREYQYWMNLCEDGTVSYFMGTQFSPGSAERSAERHLMKEL
TYKKESSCCDQAGPELSPVIGETSTLAGAVSSASYPGAAASERSAERHLMKEL
NIEINTEQTPPAAPAIASFQQLGKOYSFEDPSSALYELLVEYISEYKAPT
GLPGYKIVLDFSPKENGSDIVHYAVTFNGEALNTWDLISLHKNVENHGLVE
LDDKPTAVYITSSILNDYIAETLHONFLMGNSSLPDPKSLQILNVRGVLVLPQEEIWM
NQSSSLQVYITSSILNDYIAETLHONFLMGNSSLPDPKSLQILNVRGVLVLPQEEIWM
SYLGDIVSPKLASPKSVLSSPEVLGSSLLHSYPAVLQIDLPVAPGRTSGSS
ILEDNTSESDVSDLPSSLIQPVPAETVPPMEDSDMLITSSPFLTSSVIEDLA
KIDTPSGDLSASVSDKLVSPFPDTSVEKEFIFESGLSGSKNDVIDPWPSE
TLEKTTEPLSSWSQEDTLPTESIEKLHMYFTQMIPEPSAHRYGDDGPIYETEEES
HYRSTPIPAESATQPTLSISKHTSDVPDIDSYTKAPFLATANTASTKTEDEV
NTLLKGMQVTPSPKGLDSKISVARPDMQPVMTILPESDTVMARTSSIGLSRDTL
VTPPSADRLWLKASTQPAELPPSTHSTQLEDEVMQVONISLELDQVGTDTYQPEL
TQONGKDVSYVEMPHVHTLEMPVAQPKGVTSRTQTAGALVVFSLRVNMLFS
EDLFNNKLSLEYKALEOFLELPLVQLNSLGFQNLLEILNFRNGSVNRSVFAESV
PNNVNAIYMLIEDFTCTAYQTMNLDIDRYSLDVSQDDANPCFOACNEFSCLVNP
WGEAKCKHPGLVDELPCQSCVDLPQDFCLNDKCDVMPGHCAGICRCRVGNNWY
RGHCHEFSEPEPVIGITISVLLVASAVVFLAKMLQAOVRRORPNRQPD
SLSSVENAMKYPAYSRLAGCEQYKPYSHQPFYSASEEIVGGLSREIROMYESS
DUSKEIQERMLLEYDPPFAAIVRSEHEML"

BASE COUNT 1228 a 986 c 1024 g 1130 t
ORIGIN

Query Match 0.9%; Score 38; DB 12; Length 4368;
Best Local Similarity 100.0%; Pred. No. 7.7e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3222 gccaaccttcgaagttcaggctgaatgaatttc 3259
|||||
Db 3223 gccaaccttcgaagttcaggctgaatgaatttc 3260

RESULT 9

AC026980/c
LOCUS 65674 bp DNA HTG 25-MAR-2000
Homo sapiens chromosome 21 clone RP11-65K18 map 21, LOW-PASS
SEQUENCE SAMPLING.
AC026980
VERSION 1 GI:7328857
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 65674)
AUTHORS Birren, B., Linton, J., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 21, clone RP11-65K18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 65674)
AUTHORS Birren, B., Linton, J., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, P.,
Boguslavskiy, L., Bukhgaler, B., Brown, A., Burkett, G.,
Campoliano, A., Cascie, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, A., DeArelano, K., Dewar, K., Diaz, J. S.,
Dodgson, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Gallagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

TITLE
JOURNAL
COMMENT

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehotsky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meidrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., O'Neil, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, N.
Direct Submission
Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8662
Center clone name: 65_K_18

NOTE

* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be generic and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 664: contig of 664 bp in length
* 565 764: gap of 100 bp
* 765 1456: contig of 692 bp in length
* 1457 1556: gap of 100 bp
* 1557 2235: contig of 679 bp in length
* 2236 2335: gap of 100 bp
* 2336 3002: contig of 667 bp in length
* 3003 3102: gap of 100 bp
* 3103 3789: contig of 687 bp in length
* 3790 3889: gap of 100 bp
* 3890 4593: contig of 704 bp in length
* 4594 4693: gap of 100 bp
* 4694 5378: contig of 885 bp in length
* 5379 5478: gap of 100 bp
* 5479 6167: contig of 689 bp in length
* 6168 6267: gap of 100 bp
* 6268 6942: contig of 675 bp in length
* 6943 7042: gap of 100 bp
* 7043 7711: contig of 669 bp in length
* 7712 7811: gap of 100 bp
* 7812 8488: contig of 677 bp in length
* 8489 8588: gap of 100 bp
* 8589 9285: contig of 697 bp in length
* 9286 9385: gap of 100 bp
* 9386 10069: contig of 884 bp in length
* 10070 10169: gap of 100 bp
* 10170 10845: contig of 676 bp in length
* 10846 10945: gap of 100 bp
* 10946 11632: contig of 687 bp in length
* 11633 11732: gap of 100 bp
* 11733 12421: contig of 689 bp in length
* 12422 12521: gap of 100 bp
* 12522 13206: contig of 685 bp in length
* 13207 13306: gap of 100 bp
* 13307 14002: contig of 696 bp in length

* 14003 14102: gap of 100 bp
* 14103 14782: contig of 680 bp in length
* 14783 14882: gap of 100 bp
* 14883 15569: contig of 687 bp in length
* 15570 15669: gap of 100 bp
* 15670 16357: contig of 688 bp in length
* 16358 16457: gap of 100 bp
* 16458 17144: contig of 687 bp in length
* 17145 17244: gap of 100 bp
* 17245 17919: contig of 675 bp in length
* 17920 18019: gap of 100 bp
* 18020 18690: contig of 671 bp in length
* 18691 18790: gap of 100 bp
* 18791 19479: contig of 689 bp in length
* 19480 19579: gap of 100 bp
* 19580 20273: contig of 694 bp in length
* 20274 20373: gap of 100 bp
* 20374 21048: contig of 675 bp in length
* 21049 21148: gap of 100 bp
* 21149 21798: contig of 650 bp in length
* 21799 21898: gap of 100 bp
* 21899 22592: contig of 694 bp in length
* 22593 22692: gap of 100 bp
* 22693 23373: contig of 681 bp in length
* 23374 23473: gap of 100 bp
* 23474 24157: contig of 684 bp in length
* 24158 24257: gap of 100 bp
* 24258 24935: contig of 678 bp in length
* 24936 25035: gap of 100 bp
* 25036 25715: contig of 680 bp in length
* 25716 25815: gap of 100 bp
* 25816 26501: contig of 686 bp in length
* 26502 26601: gap of 100 bp
* 26602 27303: contig of 702 bp in length
* 27304 27403: gap of 100 bp
* 27404 28099: contig of 696 bp in length
* 28100 28199: gap of 100 bp
* 28200 28895: contig of 696 bp in length
* 28896 28995: gap of 100 bp
* 28996 29700: contig of 705 bp in length
* 29701 29800: gap of 100 bp
* 29801 30496: contig of 696 bp in length
* 30497 30596: gap of 100 bp
* 30597 31297: contig of 701 bp in length
* 31298 31397: gap of 100 bp
* 31398 32097: contig of 700 bp in length
* 32098 32197: gap of 100 bp
* 32198 32890: contig of 693 bp in length
* 32891 32990: gap of 100 bp
* 32991 33673: contig of 683 bp in length
* 33674 33773: gap of 100 bp
* 33774 34457: contig of 684 bp in length
* 34458 34557: gap of 100 bp
* 34558 35244: contig of 687 bp in length
* 35245 35344: gap of 100 bp
* 35345 36030: contig of 686 bp in length
* 36031 36130: gap of 100 bp
* 36131 36797: contig of 667 bp in length
* 36798 36897: gap of 100 bp
* 36898 37575: contig of 678 bp in length
* 37576 37675: gap of 100 bp
* 37676 38338: contig of 663 bp in length
* 38339 38438: gap of 100 bp
* 38439 39135: contig of 697 bp in length
* 39136 39235: gap of 100 bp
* 39236 39928: contig of 693 bp in length
* 39929 40028: gap of 100 bp
* 40029 40700: contig of 672 bp in length
* 40701 40800: gap of 100 bp
* 40801 41499: contig of 699 bp in length
* 41500 41599: gap of 100 bp
* 41600 42152: contig of 553 bp in length
* 42153 42252: gap of 100 bp

* 42253 42927: contig of 675 bp in length
* 42928 43027: gap of 100 bp
* 43028 43706: contig of 679 bp in length
* 43707 43806: gap of 100 bp
* 43807 44477: contig of 671 bp in length
* 44478 44577: gap of 100 bp
* 44578 45245: contig of 668 bp in length
* 45246 45345: gap of 100 bp
* 45346 46045: contig of 700 bp in length
* 46046 46145: gap of 100 bp
* 46146 46823: contig of 678 bp in length
* 46824 46923: gap of 100 bp
* 46924 47581: contig of 658 bp in length
* 47582 47681: gap of 100 bp
* 47682 48372: contig of 691 bp in length
* 48373 48472: gap of 100 bp
* 48473 49144: contig of 672 bp in length
* 49145 49244: gap of 100 bp
* 49245 49929: contig of 685 bp in length
* 49930 50029: gap of 100 bp
* 50030 50724: contig of 695 bp in length
* 50725 50824: gap of 100 bp
* 50825 51510: contig of 686 bp in length
* 51511 51610: gap of 100 bp
* 51611 52297: contig of 687 bp in length
* 52298 52397: gap of 100 bp
* 52398 53085: contig of 688 bp in length
* 53086 53185: gap of 100 bp
* 53186 53868: contig of 683 bp in length
* 53869 53968: gap of 100 bp

Query Match 0.5%; Score 22; DB 68; Length 65674;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 tatcacaccttttgggaagcag 950
|||||
Db 37255 TATCACCTTTTGGGAGCAG 37234

RESULT 10

HS511B24

LOCUS

DEFINITION Homo sapiens DNA sequence from clone 511B24 on chromosome 2q11.2-12. Contains the TOP1 gene for Topoisomerase I, the PLCGL gene for 1-Phosphatidylinositol-4,5-Bisphosphate Phosphodiesterase Gamma 1 (EC 3.1.4.11, PLC-Gamma-1, Phospholipase C-Gamma-1 PLC-II, PLC-148), the KIAA0395 gene for a probable Zinc Finger Homeobox protein and a 60S Ribosomal Protein L23 LIKE pseudogene. Contains a predicted CpG island, ESTs, STSs and GSSs, complete sequence.
AL022394

ACCESSION

VERSION

KEYWORDS

HTG; 1-Phosphatidylinositol-4,5-Bisphosphate Phosphodiesterase Gamma 1; 60S Ribosomal Protein L23; Phospholipase C-Gamma-1; PLC-148; PLC-Gamma-1; PLC-II; PLCGL; TOP1; Topoisomerase I. human.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 146547)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (10-JUL-1998) E-mail enquiries: humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jul 4, 1998 this sequence version replaced gi:3281979.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence is the entire insert of clone 511B24. This sequence has been finished according to sequence map criteria as follows. An attempt is made

to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

511B24 is from the library RCI3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

FEATURES

```

source
    1. .146547
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="20"
        /map="q11.2-12"
        /clone="RP3-511B24"
        /clone_11b="RPCI-3"
        /note="AluX repeat: matches 301. .1 of consensus"
        /complement(1348. .1462)
        /note="L1 repeat: matches 4845. .4733 of consensus"
        1555. .1632
        /note="2 copies 39 mer 89% conserved"
        /complement(3104. .3165)
        /note="AluJb repeat: matches 66. .1 of consensus"
        4613. .4903
        /note="L1 repeat: matches 5105. .5390 of consensus"
        4745. .4950
        /note="L1ME3A repeat: matches 1. .198 of consensus"
        4946. .5172
        /note="L1ME1 repeat: matches 375. .614 of consensus"
        5180. .5475
        /note="AluX repeat: matches 1. .298 of consensus"
        /complement(5387. .5998)
        /note="match: GSS B52960"
        5479. .5613
        /note="L1ME2 repeat: matches 627. .763 of consensus"
        /complement(6779. .6823)
        /note="AluJ repeat: matches 296. .250 of consensus"
        /complement(7748. .7892)
        /note="AluX repeat: matches 275. .135 of consensus"
        /join(<7958. .8101,9146> .9331,10547 .10615,11454 .11568,
        13345. .13472,16872. .16966,17182. .17331,18371. .19663)
        /gene="TOP1"
        /note="match: CDNAS J03250 X16479 U07804 D10061 Z21624
        M27913 U07806 Z21625 L20632 M55632 AB000407 L07777 match:
        ESTs AA655095 AA871201 AA623645 D55538 D54890 AA475125
        T85685 F07589 R60159 AA848218 AA587119 AA693724 R74888
        AA267891 R20925 AA814104 AA627288 C86359 AA189982 R32386
        AA438250 AA216922 AA260815 AA203865 AA371544 AA830283
        AA836116 AA285998 AA232856 AA794240 AA144395 AA345194
        AA170792 AA792988 AA655255 AA148456 R91594 AA513865 R63040
        H94352 R00391 N39131 AA663285 AA708654 AA267357 N46540
        AA826178 AA977503 R69110 N83271 AA745864 R68389 T85685
        AA233029 AA337101 AA565848 AA827288 AA501195 H65749 R38217
        R60159 AA809301 AA914511 AA769088 AA474578 R76622 AA933044
        AA574432 AA574450 AA573154 AA747450"
        /evidence=not_experimental
        /product="dJ511B24.1 (Topoisomerase I)"
        7958. .19663
        /gene="TOP1"
        /join(<7958. .8101,9146> .9331,10547 .10615,11454. .11568,
        13345. .13472,16872. .16966,17182. .17331,18371. .18473)
        /gene="TOP1"
        /note="match: Proteins Q07050 O43256 P79994 P14512 Q04750
        P1387 O13463 Q00313 Q94705 P07799 P04786 CE12282 CE12284
        P30189"
        /codon_start=1
        /evidence=not_experimental
        /product="dJ511B24.1 (Topoisomerase I)"
        /protein_id="CAA18536.1"

```

```

/db_xref="GI:3336978"
/db_xref="SPTREMBL:Q12855"
/translation="GEXDKQKYEYARRLKKCVDKIRNOYREDWKSKEMKVRQAVALLY
FIDKLALRAGNEKEGETADYVGCSSLVEHINLHPDLGGQYVVEFFLKGDSIRYY
NKVPEKVFKNLQIFMENKOPEDDLFRNLNTGILNKHLDLMEGLTAKVFRTYNASI
TLQOLKELTAPDENIPAKILSYNRRANRAVAILCNHQRAPPKTFKSMNLTQTKIDAK
KEQLADARDLKSADAKVMKDARTKKVSKKKAVQRLKEQLKLEQVQATDREENT
QIALGTSKLNLDPRITVAVCKKGVPIEKIYNTQREKFAWIDMAEDVEF"
8772. .9075
/note="AluY repeat: matches 1. .301 of consensus"
9455. .9620
/note="FAM repeat: matches 12. .175 of consensus"
10107. .10207
/note="MER42c repeat: matches 1235. .1345 of consensus"
10208. .10359
/note="4 copies 38 mer 83% conserved"
10396. .10496
/note="MER42c repeat: matches 1430. .1532 of consensus"
10993. .11245
/note="AluJb repeat: matches 1. .280 of consensus"
12247. .12527
/note="AluJo repeat: matches 1. .299 of consensus"
/complement(12562. .12860)
/note="AluY repeat: matches 301. .1 of consensus"
14510. .14642
/note="AluJb repeat: matches 2. .134 of consensus"
/complement(15919. .16215)
/note="AluY repeat: matches 298. .3 of consensus"
18023. .18254
/note="MIR repeat: matches 2. .236 of consensus"
19638. .19643
/gene="TOP1"
19657
/gene="TOP1"
19663
/gene="TOP1"
20127. .20230
/note="match: GSS AQ016533"
/complement(21628. .21830)
/note="L1MB8 repeat: matches 878. .645 of consensus"
22553. .22588
/note="18 copies 2 mer tg 100% conserved"
/complement(22602. .22875)
/note="AluSg repeat: matches 298. .1 of consensus"
22906. .23047
/note="AluJo repeat: matches 137. .281 of consensus"
/complement(23273. .24163)
/note="L1PA2 repeat: matches 891. .1 of consensus"
/complement(24014. .24379)
/note="L1 repeat: matches 5390. .5025 of consensus"
24365. .24952
/note="L1 repeat: matches 4432. .5017 of consensus"
/complement(24953. .25251)
/note="AluX repeat: matches 298. .1 of consensus"
/complement(25290. .25589)
/note="AluJo repeat: matches 295. .2 of consensus"
25774. .26077
/note="L1ME2 repeat: matches 586. .901 of consensus"
26124. .26418
/note="AluSg repeat: matches 1. .297 of consensus"
27555. .27847
/note="AluSg repeat: matches 1. .300 of consensus"
28720. .28847
/note="AluSp repeat: matches 172. .302 of consensus"
28929. .28977
/note="MIR repeat: matches 20. .68 of consensus"
28981. .29073
/note="MIR repeat: matches 111. .208 of consensus"
/complement(29082. .29127)
/note="MIR repeat: matches 241. .193 of consensus"
30351. .30499
/note="MIR repeat: matches 34. .190 of consensus"
31265. .31313

```


ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 171741)
Parker/A.
Direct Submission
Submitted (20-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 21, 2000 this sequence version replaced gi:7024299.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj354M18
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 161469 bases at least Q40
Consensus quality: 165558 bases at least Q30
Consensus quality: 168131 bases at least Q20
Insert size: 170341; sum-of-contigs
Insert size: 160326; 10.6% error; agarose-fp
Quality coverage: 3.42x in Q20 bases; sum-of-contigs Quality
coverage: 3.63x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently * consists
of 16 contigs. The true order of the pieces is * not known and
their order in this sequence record is * arbitrary. Where the
contigs adjacent to the vector can * be identified, they are
labelled with 'clone_end' in the * feature table. Some order and
orientation information * can tentatively be deduced from paired
sequencing reads * which have been identified to span the gap
between two * contigs. These are labelled as part of the same *
'fragment_chain', and the order and relative orientation * of the
pieces within a fragment_chain is reflected in * this file. Gaps
between the contigs are represented as * runs of N, but the exact
sizes of the gaps are unknown. * This record will be updated with
the finished sequence as * soon as it is available and the
accession number will be * preserved.
* 1 14050 contig of 14050 bp in length; fragment_chain 1
* 14151 47417 contig of 33267 bp in length; fragment_chain 1
* 47518 51692 contig of 4175 bp in length; fragment_chain 1 *
51793 65369 contig of 13577 bp in length; fragment_chain 2 *
65470 68343 contig of 2874 bp in length; fragment_chain 2 *
68444 79945 contig of 11503 bp in length; fragment_chain 2 *
80047 103635 contig of 23589 bp in length; fragment_chain 2 *
103736 107122 contig of 3387 bp in length; fragment_chain 2 *
107223 117802 contig of 10580 bp in length; fragment_chain 3 *
117903 124666 contig of 6764 bp in length; fragment_chain 3 *
124767 126003 contig of 1237 bp in length; fragment_chain 3 *
* 126104 147223 contig of 21120 bp in length
* 147324 160327 contig of 13004 bp in length
* 160428 161556 contig of 1129 bp in length
* 161657 162951 contig of 1295 bp in length
* 163052 171741 contig of 8690 bp in length.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1. .171741
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-354M18"
/clone_lib="RPC1-3"
1. .14050
/note="assembly_fragment:00804
fragment_chain:1
misc_feature

clone_end:T7
vector_side:left
14151..47417
/note="assembly_fragment:00210
fragment_chain:1"
47518..51692
/note="assembly_fragment:00683
fragment_chain:1"
51793..55369
/note="assembly_fragment:01494
fragment_chain:2"
65470..68343
/note="assembly_fragment:00580
fragment_chain:2"
68444..79946
/note="assembly_fragment:00824
fragment_chain:2"
80047..103635
/note="assembly_fragment:00744
fragment_chain:2"
103736..107122
/note="assembly_fragment:00799
fragment_chain:2"
107223..117802
/note="assembly_fragment:00472
fragment_chain:3"
117903..124666
/note="assembly_fragment:00532
fragment_chain:3"
124767..126003
/note="assembly_fragment:00300"
126104..147223
/note="assembly_fragment:00304"
147324..160327
/note="assembly_fragment:01083"
160428..161556
/note="assembly_fragment:01271"
161657..162951
/note="assembly_fragment:01365"
163052..171741
/note="assembly_fragment:01310
clone_end:SP6
vector_side:right"
BASE COUNT 54119 a 31984 c 32447 g 51684 t 1507 others
ORIGIN
Query Match 0.5%; Score 22; DB 40; Length 171741;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4045 atagggcacactgtttttttt 4066
|||||
Db 128398 ATAGGCACACGTGTTTTTTT 128377
RESULT 13
HS134E15/c
LOCUS
DEFINITION
HS134E15 177163 bp DNA PRI 23-NOV-1999
Human DNA sequence from clone 134E15 on chromosome 6q21. Contains
Blimp-1, apoptosis specific protein similar to yeast APG5 ESTs,
GSSs and retroviral sequence, complete sequence.
AL022067
VERSION AL022067.1 GI:3395491
KEYWORDS RTG; Blimp-1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 177163)
Tubby.B.
Direct Submission
Submitted (19-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk
JOURNAL

Clone requests: clonerequest@sanger.ac.uk
 On Aug 5, 1998 this sequence version replaced gi:3355451.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence is the entire insert of clone 134E15. This sequence
 has been finished according to sequence map criteria as follows. An
 attempt is made to resolve all sequencing problems, such as
 compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 This sequence was generated from part of bacterial clone contigs of
 human chromosome 6, constructed by the Sanger Centre Chromosome 6
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 134E15 is from the library RPC11 constructed at the Roswell Park
 Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/VECTOR:PCYAC2>.

FEATURES

source

```

location/Qualifiers
  1..177163
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /map="q21"
    /clone_lib="RPC1-1"
  repeat_region
    1..94
      /note="AluJo repeat: matches 194..287 of consensus"
      join(800..921,3764..3865,4485..4737,10010..11118,
11556..11684,12096..15114)
      /gene="dJ134E15.1"
      /evidence=not_experimental
      800..15114
        /gene="dJ134E15.1"
        join(<802..921,3766..3867,4487..4739,10012..11120,
11558..11686,12098..12673)
        /gene="dJ134E15.1"
        /codon_start=1
        /evidence=not_experimental
        /product="dJ134E15.1 (B1imp-1)"
        /protein_id="CAB36862.1"
        /db_xref="GI:4455442"
        /db_xref="SPTREMBL:O95914"
        /translation="VIGMSKEYIPKTRFGPLIGIYNTDVPKNARKYFWRESAS
LDMFKRSDFSVIDQIRAAARVALLWQISRGELHHFIDGFNEEKSNWYVNPAP
HSPRQNLAAACONGNIVYTIKPIIPANQELLLWYCRDFAERLHYRYPGELTWMNLTQ
TQSSLKQSTENELCPKNVPKREYSVKELKLDNSPKGKDLRYNSIPLTSEKLDL
DFRRGSPMPYIPRVYIPRALPDEDFUKASLAGIERTYITRPIFSSTTPSPSA
RSSPQSLKSSPSHSPGTVSPVPGSQEHRDSYAYLNASTYEGTGLGSPYAPLPH
LPAPITPSYNAHYRFLPLPPYGMNCLGSAVSMGNNFGLPRLCPYVSNLGGGS
LPHMLNPLTSPSSLPDQARLLQPEHREVLVPAPHSFPTGAASMKDKACSPY
SGSPAGTAAATAEHVQPKATSAAMAAPSSDEAMNLIKRNNTGVTPLPYLKKONG
KIKYCNVCATKFGQLSNLKVHLRVHSGERPCKQCNGFTQLAHLQRYLHVHTEK
PHEQCQVCHRFSSTNLKTLHLRSGEKYQCKVCPAKTFQFVHLKRLKRLHREDFH
KCSQCKNLIHLCSLKXNLYLGNCAAPAFPLEDLTRINEETKEFDISDNADRLDEY
EDDISVSVVEREILAVRKEETGLKVSLORMNGNLLSGCSLYESSDLPMLKLP
PSNPLVPVKYQETVEPMDP"
        complement(1335..1632)
      /note="AluSc repeat: matches 299..1 of consensus"
      complement(5591..5709)
      /note="MER42c repeat: matches 1538..1420 of consensus"
      complement(7667..7726)
      /note="MIR repeat: matches 258..199 of consensus"
      complement(9396..9696)
      /note="AluSp repeat: matches 303..2 of consensus"
      complement(9792..9830)
      /note="MIR repeat: matches 158..120 of consensus"
      11241..11309
        /note="MIR repeat: matches 49..113 of consensus"

```

```

prim_transcript <12902..>13321
/gene="dJ134E15.1"
/note="match: multiple ESTs; match: AA730123 R25873"
prim_transcript 13946..14238
/gene="dJ134E15.1"
/note="match: multiple ESTs; match: 242260 AA768582"
prim_transcript <14302..>14610
/gene="dJ134E15.1"
/note="match: multiple ESTs; match: AA344205 AA372865
AA317110"
prim_transcript <14664..>15126
/note="match: multiple ESTs; match: AA827238 AA664123
AA858171 AA857721; match: AA743541 238498 T39272 AA921725
AA507413; match: N94326 AA766820"
repeat_region
/note="AluJo repeat: matches 302..1 of consensus"
18959..18951
/note="MIR repeat: matches 48..140 of consensus"
complement(20931..21229)
/note="AluSc repeat: matches 299..1 of consensus"
complement(21344..21469)
/note="MIR repeat: matches 164..36 of consensus"
23277..23411
/note="MER30 repeat: matches 2..142 of consensus"
23412..23704
/note="AluSg repeat: matches 1..292 of consensus"
23705..23801
/note="MER30 repeat: matches 133..230 of consensus"
complement(24029..24215)
/note="L1MB2 repeat: matches 915..726 of consensus"
complement(25622..25713)
/note="MIR repeat: matches 192..97 of consensus"
26697..26985
/note="AluSx repeat: matches 1..289 of consensus"
27364..27446
/note="MIR repeat: matches 47..142 of consensus"
complement(29028..29070)
/note="MIR2 repeat: matches 145..104 of consensus"
29179..29383
/note="L1MB3A repeat: matches 411..613 of consensus"
complement(30187..30302)
/note="MIR repeat: matches 146..20 of consensus"
30808..31092
/note="AluSx repeat: matches 1..298 of consensus"
32479..32776
/note="AluSx repeat: matches 1..298 of consensus"
33235..33537
/note="AluSg repeat: matches 1..303 of consensus"
33748..34038
/note="AluSx repeat: matches 1..291 of consensus"
complement(35979..36266)
/note="AluSx repeat: matches 302..1 of consensus"
36635..36944
/note="AluJo repeat: matches 1..302 of consensus"
37441..37512
/note="MIR2 repeat: matches 71..144 of consensus"
complement(38479..38551)
/note="MER33 repeat: matches 316..242 of consensus"
38555..38848
/note="AluSx repeat: matches 10..302 of consensus"
complement(38856..38963)
/note="MER33 repeat: matches 254..146 of consensus"
38965..39240
/note="AluSg repeat: matches 1..292 of consensus"
complement(39254..39338)
/note="MER33 repeat: matches 158..79 of consensus"
41730..42041
/note="AluY repeat: matches 1..301 of consensus"
42625..42754
/note="MIR repeat: matches 90..225 of consensus"
43357..43473
/note="MIR2 repeat: matches 2..121 of consensus"
complement(44420..44720)

```

```
/note="AluSg repeat: matches 299. .1 of consensus"
46124. .46410
/note="AluSg repeat: matches 1. .300 of consensus"
complement(46156. .46698)
/note="MIR repeat: matches 185. .3 of consensus"
47373. .47502
/note="MIR repeat: matches 54. .187 of consensus"
48336. .48548
/note="MIR repeat: matches 13. .258 of consensus"
complement(49685. .49976)
/note="AluSg repeat: matches 301. .3 of consensus"
50183. .50269
/note="MIR repeat: matches 65. .151 of consensus"
50566. .50725
/note="MIR repeat: matches 56. .207 of consensus"
complement(51530. .51932)
/note="match: GSS A0016620 clone 230F8"
complement(52626. .52730)
/note="MIR repeat: matches 153. .46 of consensus"
55378. .55510
/note="FLAM_C repeat: matches 1. .133 of consensus"
complement(55728. .56028)
/note="AluSg repeat: matches 300. .1 of consensus"
56225. .56502
/note="AluSx repeat: matches 3. .280 of consensus"
complement(56504. .56882)
/note="LLMB6 repeat: matches 847. .483 of consensus"
56882. .57244
/note="THE1B repeat: matches 1. .364 of consensus"
57245. .57330
/note="THE1B-INTERNAL repeat: matches 1. .86 of consensus"
57330. .57697
/note="THE1B-INTERNAL repeat: matches 1212. .1580 of
consensus"
57698. .58052
/note="THE1B repeat: matches 1. .352 of consensus"
complement(58034. .58225)
/note="LLMB6 repeat: matches 510. .320 of consensus"
complement(58793. .59097)

Query Match      0.5%; Score 22; DB 10; Length 177163;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4045 atagggcacactgtttttttt 4066
|||||
Db 134768 ATAGGCACACTGTTTTTTT 134747

RESULT 14
AC024189 AC024189 193756 bp DNA HTG 10-MAR-2000
LOCUS Homo sapiens clone RP11-425G1, *** SEQUENCING IN PROGRESS ***, 48
DEFINITION unordered pieces.
ACCESSION AC024189
VERSION AC024189.2 GI:7212930
KEYWORDS HTG; HTGS-PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 193756)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193756)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 10, 2000 this sequence version replaced gi:7105562.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_MH0425G01
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1283: contig of 1283 bp in length
* 1284 1401: gap of unknown length
* 1402 2590: contig of 1189 bp in length
* 2591 2708: gap of unknown length
* 2709 4252: contig of 1544 bp in length
* 4253 4370: gap of unknown length
* 4371 5888: contig of 1518 bp in length
* 5889 6007: gap of unknown length
* 6007 8317: gap of unknown length
* 8318 9659: contig of 1342 bp in length
* 9660 9777: gap of unknown length
* 9778 10880: contig of 1103 bp in length
* 10881 10998: gap of unknown length
* 10999 12262: contig of 1264 bp in length
* 12263 12380: gap of unknown length
* 12381 13974: contig of 1594 bp in length
* 13975 14092: gap of unknown length
* 14093 15236: contig of 1144 bp in length
* 15237 15354: gap of unknown length
* 15355 17382: contig of 2028 bp in length
* 17383 17500: gap of unknown length
* 17501 19086: contig of 1586 bp in length
* 19087 19204: gap of unknown length
* 19205 20419: contig of 1215 bp in length
* 20420 20537: gap of unknown length
* 20538 23008: contig of 2471 bp in length
* 23009 23126: gap of unknown length
* 23127 25277: contig of 2151 bp in length
* 25278 25395: gap of unknown length
* 25396 26994: contig of 1599 bp in length
* 26995 27112: gap of unknown length
* 27113 30660: contig of 3548 bp in length
* 30661 30778: gap of unknown length
* 30779 32930: contig of 2152 bp in length
* 32931 33048: gap of unknown length
* 33049 34205: contig of 1157 bp in length
* 34206 34323: gap of unknown length
* 34324 36309: contig of 1986 bp in length
* 36310 36427: gap of unknown length
* 36428 39611: contig of 3184 bp in length
* 39612 39729: gap of unknown length
* 39730 42507: contig of 2778 bp in length
* 42508 42625: gap of unknown length
* 42626 45319: contig of 2694 bp in length
* 45320 45437: gap of unknown length
* 45438 47435: contig of 1998 bp in length
* 47436 47553: gap of unknown length
* 47554 51380: contig of 3827 bp in length
* 51381 51498: gap of unknown length
* 51499 54618: contig of 3120 bp in length
* 54619 54736: gap of unknown length
* 54737 58395: contig of 3659 bp in length
* 58396 58513: gap of unknown length
* 58514 63042: contig of 4529 bp in length
* 63043 63160: gap of unknown length
* 63161 66889: contig of 3729 bp in length
* 66890 67007: gap of unknown length
```

* 67008 71803: contig of 4796 bp in length
* 71804 71921: gap of unknown length
* 71922 75709: contig of 3788 bp in length
* 75710 75827: gap of unknown length
* 75828 80632: contig of 4805 bp in length
* 80633 80750: gap of unknown length
* 80751 85570: contig of 4820 bp in length
* 85571 85688: gap of unknown length
* 85689 91860: contig of 6172 bp in length
* 91861 91978: gap of unknown length
* 91979 97454: contig of 5476 bp in length
* 97455 97572: gap of unknown length
* 97573 103682: contig of 6110 bp in length
* 103683 108800: gap of unknown length
* 108801 108671: contig of 4871 bp in length
* 108672 108789: gap of unknown length
* 108790 113172: contig of 4383 bp in length
* 113173 113290: gap of unknown length
* 113291 119772: contig of 6482 bp in length
* 119773 119890: gap of unknown length
* 119891 125034: contig of 5144 bp in length
* 125035 125152: gap of unknown length
* 125153 130323: contig of 5171 bp in length
* 130324 130441: gap of unknown length
* 130442 137283: contig of 6842 bp in length
* 137284 137401: gap of unknown length
* 137402 143236: contig of 5835 bp in length
* 143237 143354: gap of unknown length
* 143355 151101: contig of 7747 bp in length
* 151102 151219: gap of unknown length
* 151220 161052: contig of 9833 bp in length
* 161053 161170: gap of unknown length
* 161171 172344: contig of 11174 bp in length
* 172345 172462: gap of unknown length
* 172463 183308: contig of 10846 bp in length
* 183309 183425: gap of unknown length
* 183426 193756: contig of 10331 bp in length.

FEATURES

source

1. 193756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-425G1"

BASE COUNT 46718 a 46890 c 46444 g 48091 t 5613 others

ORIGIN

Query Match 0.5%; Score 22; DB 56; Length 193756;
Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 tatccaccttttggggaagcag 950

Db 191480 TATCCACCTTTTGGGAAGCAG 191501

RESULT 15
AL133509/c

LOCUS AL133509 205746 bp DNA HTG 10-FEB-2000
DEFINITION Homo sapiens chromosome 6 clone RP3-474G24, *** SEQUENCING IN
PROGRESS ***, 20 unordered pieces.

ACCESSION AL133509

VERSION AL133509.6 GI:6983041

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 205746)

REFERENCE

AUTHORS Parker, A.

TITLE Direct Submission

JOURNAL Submitted (10-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

COMMENT

On Feb 16, 2000 this sequence version replaced gi:6967355.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00035 Length: 1065bp
Contig_ID: 00507 Length: 2401bp
Contig_ID: 00575 Length: 2384bp
Contig_ID: 00656 Length: 1250bp
Contig_ID: 00942 Length: 2757bp
Contig_ID: 00965 Length: 1073bp
Contig_ID: 00967 Length: 1485bp
Contig_ID: 01072 Length: 2169bp
Contig_ID: 01208 Length: 3773bp
Contig_ID: 01385 Length: 2695bp
Contig_ID: 01554 Length: 4184bp
Contig_ID: 01641 Length: 1135bp
Contig_ID: 01938 Length: 3041bp
Contig_ID: 02405 Length: 1001bp
Contig_ID: 02466 Length: 1154bp
Contig_ID: 02760 Length: 4146bp
Contig_ID: 03304 Length: 150691bp
Contig_ID: 03507 Length: 2142bp
Contig_ID: 03738 Length: 1000bp
Contig_ID: 03739 Length: 1000bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1065: contig of 1065 bp in length
1066 1865: gap of 800 bp
1866 4266: contig of 2401 bp in length
4267 5066: gap of 800 bp
5067 7450: contig of 2384 bp in length
7451 8250: gap of 800 bp
8251 9500: contig of 1250 bp in length
9501 10300: gap of 800 bp
10301 13057: contig of 2757 bp in length
13058 13857: gap of 800 bp
13858 14930: contig of 1073 bp in length
14931 15730: gap of 800 bp
15731 17215: contig of 1485 bp in length
17216 18015: gap of 800 bp
18016 20184: contig of 2169 bp in length
20185 20984: gap of 800 bp
20985 24757: contig of 3773 bp in length
24758 25557: gap of 800 bp
25558 28252: contig of 2695 bp in length
28253 29052: gap of 800 bp
29053 33236: contig of 4184 bp in length
33237 34036: gap of 800 bp
34037 35171: contig of 1135 bp in length
35172 35971: gap of 800 bp
35972 39012: contig of 3041 bp in length
39013 39812: gap of 800 bp
39813 40813: contig of 1001 bp in length
40814 41613: gap of 800 bp
41614 42767: contig of 1154 bp in length
42768 43567: gap of 800 bp
43568 47713: contig of 4146 bp in length
47714 48513: gap of 800 bp
48514 199204: contig of 150691 bp in length
199205 200004: gap of 800 bp
200005 202146: contig of 2142 bp in length
202147 202946: gap of 800 bp
202947 203946: contig of 1000 bp in length
203947 204746: gap of 800 bp

* 204747 205746: contig of 1000 bp in length.

FEATURES
source

1..205746
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-474G24"
/clone_lib="RPCI-3"

BASE COUNT 61380 a 38226 c 36339 g 54598 t 15203 others
ORIGIN

Query Match

Best Local Similarity 0.5%; Score 22; DB 32; Length 205746;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4045 atagggcacactgtttttttt 4066

|||||

Db 133361 ATAGGGCACACTGTTTTTTT 133340

Search completed: September 3, 2000, 17:19:46
Job time: 24903 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 12:24:55 ; Search time 7000.17 Seconds
(without alignments)
831.393 Million cell updates/sec

Title: US-09-183-972-3
Perfect score: 3261
Sequence: 1 taacacgaaggttatcct.....tactatatgacataatcaat 3261

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_bal.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: em_fun.*
17: em_hum1.*
18: em_hum2.*
19: em_in.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_sy.*
29: em_un.*
30: em_v1.*
31: gb_htg1.*
32: gb_htg2.*
33: gb_in1.*
34: gb_in2.*
35: em_bal.*
36: em_ba2.*
37: em_hum3.*
38: em_hum4.*
39: gb_pr4.*
40: gb_htg3.*
41: gb_htg4.*
42: gb_htg5.*
43: gb_htg6.*

44: gb_htg7.*
45: em_htg1.*
46: em_htg2.*
47: em_htg3.*
48: em_hum5.*
49: gb_pl3.*
50: gb_pr5.*
51: gb_htg8.*
52: gb_htg9.*
53: gb_htg10.*
54: gb_htg11.*
55: gb_htg12.*
56: gb_htg13.*
57: gb_htg14.*
58: gb_in3.*
59: gb_htg15.*
60: gb_htg16.*
61: gb_htg17.*
62: em_htg4.*
63: em_htg5.*
64: em_htg6.*
65: em_htg7.*
66: em_hum6.*
67: gb_htg18.*
68: gb_htg19.*
69: gb_htg20.*
70: gb_htg21.*
71: gb_htg22.*
72: gb_htg23.*
73: gb_htg24.*
74: gb_htg25.*
75: gb_htg26.*
76: gb_htg27.*
77: gb_htg28.*
78: gb_htg29.*
79: gb_htg30.*
80: gb_htg31.*
81: gb_v1.*
82: gb_v12.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3196.6	98.0	3268	39	AF047492 Homo sapi
2	814	25.0	1235	11	HSIMPG17
3	535.8	16.4	816	11	HSIMPG13
4	535.8	16.4	194704	32	AL157379 Homo sapi
5	511	15.7	555	11	AF047491 Macaca fa
6	249	7.6	444	11	HSIMPG10
7	237.4	7.3	537	11	HSIMPG02
8	237.4	7.3	194704	32	AL157379 Homo sapi
9	226	6.9	448	11	HSIMPG14
10	203.4	6.2	3989	50	AF157624 Homo sapi
11	203.4	6.2	4165	39	AF173155 Homo sapi
12	200.4	6.1	4368	12	RNU76717 Rattus norv
13	189.2	5.8	477	11	HSIMPG15
14	174.4	5.3	310	11	HSIMPG01
15	169.8	5.2	422	11	HSIMPG03
16	142.4	4.4	377	11	HSIMPG07
17	115.6	3.5	311	11	HSIMPG06
18	103	3.2	178064	54	AC023311 Homo sapi
19	81.4	2.5	438	11	HSIMPG12
20	79	2.4	347	11	HSIMPG11
21	69.4	2.1	304	11	HSIMPG05
22	65.4	2.0	300	11	HSIMPG08
23	61.8	1.9	380	11	HSIMPG16
24	57.2	1.8	7218	5	I66494 Sequence 14

```

c 25 54.8 1.7 247659 79 AC010172 Homo sapi
c 26 53.8 1.6 73020 51 AC022851 Homo sapi
c 27 52.8 1.6 68746 43 AC021008 Mus muscu
c 28 52.2 1.6 70006 31 AC022671 Homo sapi
c 29 52.2 1.6 22448 31 PFNALP4
c 30 51 1.6 83440 53 AC024285
c 31 50.8 1.6 975 33 DD087514
c 32 50.8 1.6 61864 33 CEY50E8A
c 33 50.6 1.6 128609 40 CNS05TC1
c 34 50.6 1.6 129404 51 AC013349 Homo sapi
c 35 50.4 1.5 253307 33 PFNAL3P7
c 36 50.2 1.5 165715 56 AC024116
c 37 50 1.5 65813 75 AC046180 Homo sapi
c 38 50 1.5 68879 72 AC031982
c 39 49.8 1.5 91949 54 AC021254
c 40 49.6 1.5 5954 14 CLOCLSB
c 41 49.6 1.5 6243 14 SVU90555
c 42 49.6 1.5 6588 14 CLOCLSA
c 43 49.6 1.5 6857 14 SVU90554
c 44 49.6 1.5 7358 14 M77169 Cloning vec
c 45 49.6 1.5 10206 1 PIP404CG M32882 Plasmid pip

```

ALIGNMENTS

```

RESULT 1
AF047492
LOCUS AF047492 3268 bp mRNA PRI 26-OCT-1999
DEFINITION Homo sapiens interphotoreceptor matrix proteoglycan 150 (IMPGL)
ACCESSION AF047492
VERSION AF047492.2 GI:6118565
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3268)
AUTHORS Kuehn,M.H. and Hageman,G.S.
TITLE Expression and characterization of the IPM 150 gene (IMPGL)
product, a novel human photoreceptor cell-associated
chondroitin-sulfate proteoglycan
JOURNAL Matrix Biol. 18 (5), 509-518 (1999)
MEDLINE 20068045
REFERENCE 2 (bases 1 to 3268)
AUTHORS Kuehn,M.H. and Hageman,G.S.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1998) Ophthalmology and Visual Sciences,
University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA
52240, USA
REFERENCE 3 (bases 1 to 3268)
AUTHORS Kuehn,M.H. and Hageman,G.S.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1999) Ophthalmology and Visual Sciences,
University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA
52240, USA
REMARK Sequence update by submitter
COMMENT On Oct 26, 1999 this sequence version replaced gi:2906231.
FEATURES
source
1..3268
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="between markers D6S280 and D6S1659"
/tissue_type="retina"
1..3268
/gene="IMPGL"
131..2524
/gene="IMPGL"
/note="IPM 150"
/codon_start=1
/product="interphotoreceptor matrix proteoglycan 150"

```

```

/protein_id="AAC03789.2"
/db_xref="GI:6118566"
/translation="MYLETRRAIFVFWIFLQVQCKTIDISINIHSETKIDIDNPPRNET
TESTEKMYKMTMRRIFDLAKHRTKRSFAFFPTGVKVCPOESMKOILDSLOAYVRLVC
QEAWEAYRFLDRIPDTGEQDWNVSICQETFCLEFDIGKNFSQSHLDLQORIKQ
RSFPDRKDEISAETLGEPTGIVISDVANVSLGPPFTDPDDLLEILLNLTNDTKQ
LPGFKIHVLGFRPKKDKGSSSEMQLTAIFKRHSAEAKSPASDLSDFNQKIFKK
VYHGTMSADQPEIYLATDLKRLISKALEEQSLDVGTIOFTDEIAGSLAFGPDTO
SELTSPEDVTEATLSPELPVPEOLETVDAEHLPTDSWSPAMASTLSSEAPPF
FMASISFLSDQGTDMATQTMVLPGLTIPSTDYSAISQLALGISHPPASDDRS
SAGEDMVRHDEMDLSDTPAPSEVPELSEIVSPDHFLEDTIPVSLQYITTSMTI
APKRELVFVSLRVANNAFNSLFNKSLEVRALQOFTOLLVPLKSNITGFQKLE
ILNFRNSVIVNSKMTFAKSYNLTAKVHGVLDFEASAAQLHLEDSLSLNEPA
DQADPCFLACGEFAQCVRNTEAEACRCRPGYDSQGLDGLPGLCGPGTKCEVL
QKGCAPCLPDHSENOAYKTSVKFQNOQNKNVSKRNSELLTVEEYEFNQDMWEGN"
BASE COUNT 1044 a 724 c 564 g 836 t
ORIGIN
Query Match 98.0%; Score 3196.6; DB 39; Length 3268;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 3256; Conservative 0; Mismatches 5; Indels 7; Gaps 5;
QY 1 taacccaagaaggttatcctcaatcatctggtatcaataataattttt--ccttn 58
DB 1 TAAACCAAGAGGTTATCTCAATCATCTGTTATCAATATAATAATTTTTCACATT 60
QY 59 tgttacttttaagattgaggttctgtctgtatgtttatcagaattacc-atgcac 117
DB 61 TGTTACTTTTAAATGAGATTGAGGTGTTCTGTGATTGTTATCAGAAATACCAATGCAC 120
QY 118 aaaaaccagaatgtatttgaaactagaagactattttgttttttgattttctcca 177
DB 121 AAAACCCAGAAATGATTGGAAACTAGAGAGCTATTTTGTGTTTGGATTTTCTCCA 180
QY 178 agttcaagaaactaagatatctccattataacataaccattctgaactaaagacataga 237
DB 181 AGTTCAAGGAACCAAGATATCTCCATTAAACATATACCATTCGAAACTAAAGACATAGA 240
QY 238 caatcccccaagaataaactgaaagtactgaaaaaaatgtacaaaatgtcaactat 297
DB 241 CAATCCCCCAAGAAATGAAACAACTGAAAGTACTGAAAAAATGTACAAAATGTCAACTAT 300
QY 298 gagcagaatattcgatttggaacatcgaaacatcgaaacacacacacacacacgag 357
DB 301 GAGACGAATATTCGATTGGCAAGCATCGAACAAAAAGATCCGCATTTTCCACAGGG 360
QY 358 ggttaaagctgtccacaggaatccatgaacagatttttagacagcttctcaagcttatta 417
DB 361 GGTAAAGTCTGTCCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTTATTA 420
QY 418 tagattgagagtgtgtcagggaagcagtagtggaagcattatcgagatttctcgatcgat 477
DB 421 TAGATTGAGAGTGTCTCAGGAAGCAGTAGTGGGAAGCATATCGGATCTTCTCGATCGCAT 480
QY 478 cctcacacaggggaatatcaggactgggtcagcatctccacagcagacacattctgcct 537
DB 481 CCCTGACACAGGGGAATATCAGGACTGGGTGAGCATCTCCACAGCAGACCTCTCGCT 540
QY 538 ctttgacattgaaaaaacctcagcaattcccaggagacacctggatcttctccagcagag 597
DB 541 CTTTGACATTGGAAAAAATTCAGCAATTCACAGGAGCACCCTGGATCTTCTCCAGCAGAG 600
QY 598 aataaacagagaagtttccctgacagaaaaagataatctatgcagagaagacattggg 657
DB 601 AATAAACAGAGAAGTTTCCCTGACAGAAAAAGATGAAATATATCGCAGAGAAGACATTGGG 660
QY 658 agagcctgtgaaccattgttcatttcaacagatgttgccaacgtctcacttgggcttt 717
DB 661 AGAGCCTGTGTGAACCATTTGTCTATTTCACACAGATGTTGCAACAGCTCTACTTGGGCTTT 720
QY 718 ccctctcactcctgatgacacccctcctcactgaaattcttcgataatacactcaacgacac 777

```


Db 721 CCGTCTCACTCCCTGATGACACCTCTCTCAATGAATAATCTCGATAATCACTCAACGACAC 780
Qy 778 caagatgcttacaacagagagaaacagaaatcgctgtgtgtgagagagaggggtgga 837
Db 781 CAAGATGCCCTACACAGAGAAAGAAACAGAAATTCGCTGTGTGTGGAGGAGCAGAGGGTGA 840
Qy 838 gctcagcgtctctctgtgtaaacacagagttcaaggcagagctcgctgactccagtcgcc 897
Db 841 GCTCAGCGTCTCTCTGTGTAAACAGAGATTCAGGCGAGAGCTCGCTGACTCCAGTCCCC 900
Qy 898 atattaccaggagctagcaggaagttcccaacttcagatgcataaagatatattaaagaact 957
Db 901 ATATTACCAGGAGCTAGCAGAGAAAGTCCCACTTCAGATGCATAAAGATATTTAAGAAACT 960
Qy 958 tccagattcaaaaaatccatgtgttagatttagaccaaaagaaagaaagatggctc 1017
Db 961 TCCAGGATTCAAAAAATCCATGTGTGTAGGATTTAGACCAAAAGAAAGAAAGATGGCTC 1020
Qy 1018 aagctccacagagatgcaacttacggccatctttaagagacacagtcgacgaagcaaaaag 1077
Db 1021 AAGCTCCACAGAGATGCAACTTACGGCCATCTTTAAGAGACACAGTGCAGAGCAAAAAG 1080
Qy 1078 cccgtcaagtgcacctctgtctttttgattccaaacaaattgaaagtggaggtctatca 1137
Db 1081 CCGTCAAGTGAAGTCTCTGTCTTTGATTCCACAAAAATTTGAAAGTGAAGGATCTATCA 1140
Qy 1138 tggaaacctgagagagagacaaagcaaccagaaatctctcacagctcagacctcaaaaag 1197
Db 1141 TGGAAACATGGAGAGGAGACAAAGCAACACAGAAATCTATCTCACAGCTCAGACCTCAAAAG 1200
Qy 1198 gctgatcagcaaacgactagagaaagaaacatcttgatgtggggaacttcagttcac 1257
Db 1201 GCTGATCAGCAAGACACTAGAGNAGAACAACTTTTGGATGTGGGGACAAATTCAGTTCAC 1260
Qy 1258 tgatgaaattgctggatcactgcagcgccttggctcctgcacccaatcagagctgcccac 1317
Db 1261 TGATGAAATTTGCTGGATCACTGCCAGCTTTTGGTCTCTGACACCCAAATCAGAGCTGCCAC 1320
Qy 1318 atctttgtgttataacacagagatgctactttgagtcaggaacttctcctgttgacc 1377
Db 1321 ATCTTTTGTGTTTATAACAGAGATGCTACTTTTGTAGTCCAGAACTTCTCTGTGTGAACC 1380
Qy 1378 ccagctgagacagtgagcagagacagcagctgctcactgcacactctgggtccacc 1437
Db 1381 CCAAGCTTGAGACAGTGGACGGAGCAGAGCATGGTCTACCTTGACACTCTTGGTCTCCACC 1440
Qy 1438 tgcataggcctcactccctcctcagagctccaccttcttattggcatcaagcatctt 1497
Db 1441 TGCTATGGCCTCTACCTCCCTGTGACAGGCTCCACCTTTCTTTATGGCATCAAGCATCTT 1500
Qy 1498 ctctctgactgatcaaggccaccagatacaatggccactgaccagacaaatgctagtacc 1557
Db 1501 CTCTCTGACTGATCAAGGCCACCAGATACAAATGGCCACTGACCCAGACAAATGCTAGTACC 1560
Qy 1558 agggctcaccatccccaccagtgattattgtgcaatcagccaaactggctctgggaatttc 1617
Db 1561 AGGGCTCACCATCCCCACCACTGATTATTCTGCAATCAGCCCACTGGCTCTGGGAATTC 1620
Qy 1618 acatccactctgactctcagatgacagccgatcaagtgcagggtggcgaagatatgggtcag 1677
Db 1621 ACATCCACTGTCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGCCGAAGATATGGTCAG 1680
Qy 1678 acacctagatgaaatggatctgtcagactcctgccccactctgaggtaccagagctcag 1737
Db 1681 ACACCTTAGATGAAATGGATGCTGTGACACTCTCTGCCCCATCTGAGGTACCAGAGCTCAG 1740
Qy 1738 cgaatatgtttctgtccagatcaattcttggaggaataccactcctgtctcagatttaca 1797
Db 1741 CGAAATATGTTCTGTGCCAGATCATTTCTTGGAGGATACCACTCCTGCTCAGCTTTACA 1800
Qy 1798 gtatatcaccactagttctctatgaccattgtccccaagggccgagagctggtagttctt 1857

Db 1801 GTATATCACCACTAGTTTCTATGACCATTGCCCCCAAGGCCCGAGAGCTGGTAGTGTCTT 1860
Qy 1858 cagtcgctgtgtgtaacatggccttctccaaacgacctgttcaacaagagctctctgga 1917
Db 1861 CAGTCTGCGTGTGTAAACATGGCCTTCTCCAAACGACCTGTTTCAACAAGAGCTCTCTGA 1920
Qy 1918 gtaccgagctctggagcaacaattcacacagctgtgttccatctatctacgatccaact 1977
Db 1921 GTACCGAGCTCTGGAGCAACAATTCACACAGCTGTGTGTTCCATATCTACGATCCAATCT 1980
Qy 1978 tacagatttaagcaacttgaataacttaactcagaacggagagtgattgtaataag 2037
Db 1981 TACAGGATTTAAGCAACTTGAATACTTTAACTTCAGAACGGGAGTGTGATTGTGAATAG 2040
Qy 2038 caaatgaagttgttaagctgtgcgtataacctcaccaggtgtgtcagcgggtctt 2097
Db 2041 CAAATGAAGTTTGTGTAGTCTGTGCGGTATAACTTCACCAAGGCTGTGCACGGGGTCTT 2100
Qy 2098 ggaggtatttctgtctgcagcccaacaactccatctggaatagacagctactctct 2157
Db 2101 GGAGGATTTTGTGTCTGCTGCAGCCCAACAACCTCCATCTGGAATAGACAGCTACTCTCT 2160
Qy 2158 caacattgaaccagctgatcaagcagatccctgcgaagttcctggcctgcggcgaattgc 2217
Db 2161 CAACATTGAACCCAGCTGATCAAGCAGATCCCTGCAAGTTCTTGGCTTCGGCGAATTTGC 2220
Qy 2218 ccaatgtataaagaacacgactgaggaagcagcagcgtcgtcgtcgaacaccagatatga 2277
Db 2221 CCAATGTGTAAAGAACAGACGGACATGAGAACGGAGTGTGCTGCAAAACAGAGATATGA 2280
Qy 2278 cagccaggggagcctggagcgtctggaaccagcctctgtgg - cctggcacaagaagatg 2336
Db 2281 CAGCAGGGAGCCTGGAGGCTCTGGAACCCAGGCCCTGTGTGCCCTGGCACAAGAAATG 2340
Qy 2337 cgaggtcctccagggaaaggagctccatgc - ggtccagatcactctgaaaaa tcaagc 2394
Db 2341 CGAGTCTCTCCAGGAAAGGGAGCTCCATGCAAGTTGCCAGATCACTCTGAAAAATCAAG 2400
Qy 2395 atacaaaactagttt - aaaaagttccaaaatacaaaaatacaaaaggaatcaatcagtaaaag 2453
Db 2401 ATACAAAATCTAGTGTATAAAGTTCCAAATCAACAAAATAACAGGTAAATCAGTAAAG 2460
Qy 2454 aaattcgaattactgacctagataatgaagaaatctaaccatcaagatgtgggaagaaa 2513
Db 2461 AAATCTGAATTAATGACCGTAGAATATGAAGAAATTAACCATCAAGATTGGGAAGAAA 2520
Qy 2514 ttaaaaactgaaaatgtacaattatcacttaggctctcagagagatgatttgccttc 2573
Db 2521 TAAAAAAGTAAAAATGTACAAATTAATCTAGGCTATCTCAAGAGAGATGATTGCTTTC 2580
Qy 2574 tcaaggaataatggagacagggcaattcattgggtcaatcaaaaatacaagatacagtaaca 2633
Db 2581 TCAAGGAAAAATGGAGACAGGCATATTTCAATGGGTCTATCAAAAATCCAGACATCAGTCAACA 2640
Qy 2634 ctgagaatcagcacacacacatttcaaaatagagagatcgtacttggcaaccagca 2693
Db 2641 CTGAGAAATCAGCACACACCATATTTCAAAATATAGAGAGTCAATGTTGGCAACCACTA 2700
Qy 2694 aattctgaaaaaaagacacttacttattataaaccocccaaatgcaatcagcgaataat 2753
Db 2701 AATTCTGAAAAAAGAGACACTTACTTATTATTAACCCCAAAATGCAATCAGCGAAACAT 2760
Qy 2754 attttactattcttgatgatagtcataaataagataagcaggtttgcttccaccttc 2813
Db 2761 ATTTTACTATTCTTGGATGATAGTCAAAATGATATAAGCCAGGTTTGTCTTCCACTTC 2820
Qy 2814 cctaaaaattttactcacagatcatttgcaacaagcatagcttacttattgtttaggagac 2873
Db 2821 CCTGAAAAATTTTACTCACAGATCATTTGCAACAACATAGCTTACTTATGTGTTAGGGAC 2880
Qy 2874 tgaacaattttattgggaagcaaacctctttatgctagaaagtacatttataaagatgact 2933
Db 2881 TGAACAAATTTATGGGAAGCAAACTCTTTATATGCTAGAAAGTACATTTAAAGATGACT 2940

Db 746 TGATAGTCAAAATGATCATTAAGCCAGGTTTGCTTCCACCTTCCCTGAAAATTTTACTTCAC 805
 QY 2832 agatcatttgcacaacagctagcttactatttgggttggactgaacaattatttggaa 2891
 Db 806 AGATCATTTGCAACAGCATAGCTTACTTATTTGTTAGGACTGAACAATTTTGGAA 865
 QY 2892 gcaaaccttttatgtctagaagttacatttaaagatgactactacgcaggagatgc 2951
 Db 866 GCAAACTCTTTATATGCTAGAAAGTACATTTAAAGATGACTACTTACGCGAGGAGATGC 925
 QY 2952 aggtctcttaaacgcgatgatgtatgtgtgttaggcactgtagtgagtgatatatg 3011
 Db 926 AGTCTCTCTAAACGCATGATGTATGTAGTGTAGTGTAGTGTAGTGTATATG 985
 QY 3012 ctccacactacgtctgataaacacacacccctcagttacgtatttaggcacactagttt 3071
 Db 986 CTCCACACTACGCTGATTAACACAACTCAGTATTCAGTTATTAGGCACACTAGTTT 1045
 QY 3072 atacgcaactactgcttacatagtagactgttttggccaataatctttgaattgttct 3131
 Db 1046 ATACGCAACTACTGCTTACATAGTAGACTGTTTGTGTCCTAATAATCTTTGAATTGTTCT 1105
 QY 3132 ttaaaagaactaggttccagatacacatcacatccatggaaaattcttcttctgtact 3191
 Db 1106 TTAAGAGAACTAGGTTTCAATACATACATACCATGGAATAATCTTACTTTTGTACT 1165
 QY 3192 acacaagctattttaaagaagatgctattgttggagaaggcgaagttgactatatga 3251
 Db 1166 ACACAAGCTATTTTAAAGAAGATGCTATGTTGGAGAAAGGCGAAGTTGTACTATATGA 1225
 QY 3252 cataatcaat 3261
 Db 1226 CATAATCAAT 1235

RESULT 3
 HSIIMP13 HSIIMP13 816 bp DNA PRI 28-OCT-1998
 LOCUS Homo sapiens interphotoreceptor matrix gene (IPM150), exon 13.
 DEFINITION AF017772
 ACCESSION AF017772.1 GI:3800727
 VERSION 1
 KEYWORDS 13 of 17
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 816)
 AUTHORS Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M.,
 Schmid,M. and Weber,B.H.
 TITLE Genomic organization and chromosomal localization of the
 interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate
 for 6q-linked retinopathies
 JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
 MEDLINE 98358139
 REFERENCE 2 (bases 1 to 816)
 AUTHORS Gehrig,A., Felbor,U., Kessel,R., Hunt,D.M., Maunensee-Hussels,I.E.
 and Weber,B.H.F.
 TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)
 localized to 6q14.2-q15 in autosomal dominant Stargardt-like
 macular dystrophy, progressive bifocal choroidretinal atrophy
 (PBCRA), and North Carolina macular dystrophy (MCDRI)
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 816)
 AUTHORS Felbor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
 Hubland, Wuerzburg D-97074, Germany
 FEATURES
 1. .816
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"

(2)(a)

exon
 BASE COUNT 200 a 207 c 154 g 255 t
 ORIGIN
 Query Match 16.4%; Score 535.8; DB 11; Length 816;
 Best Local Similarity 99.6%; Pred. No. 3.4e-111;
 Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1416 ctgacactcttggctccacactgtatggcctctacactccctgtcagaagctcacctt 1475
 Db 141 CAGACACTTCTGGTCTCCACCTGCTATGGCCTCTACCTCCCTGTGAGAAGTCCACCT 200
 QY 1476 tctttatggcatcaagcattctctctgactgatacgaagcaccagatacaatggcca 1535
 Db 201 TCTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGCACCATGATAAATGGCCA 260
 QY 1536 ctgaccacacaaatgcttagtaccagggtctcaccatcccccaccagtattattctgcaatca 1595
 Db 261 CTGACCACAGCAATGCTAGTACCAGGGCTCACCATCCCACACAGTATTATTCTGCAATCA 320
 QY 1596 gccaaactgctctgggaatttccacatccacctgcattcttcagatgacagccgatcaagtcg 1655
 Db 321 GCCAACTGGCTCTGGGAATTTCCATCCACCTGCATCTTCAGATGACAGCCGATCAAGTG 380
 QY 1656 cagtggtgcgaagatagctcagacacccatagatgaatggatctgtcacactctctgcc 1715
 Db 381 CAGGTGGCGAAGATATGGTCAGACACCTAGATGAATGGATCTGTGACACTCTCTGCC 440
 QY 1716 catctgaggtaccagagctcagcgaatattgttctgtcccagatcatcttctggaggata 1775
 Db 441 CATCTGAGGTACCAGAGCTCAGCGAATATGTTCTGTCCCAGATCATTTCTTGGAGGATA 500
 QY 1776 ccaactctgtctcagctttacagtatataccacactagttcttatgaccattgcccccaagg 1835
 Db 501 CCATCTCTGTCTCAGCTTTACAGTATATCACCACCTAGTTCTATGACCATTTGCCCAAGG 560
 QY 1836 gccgagagctggttagttcttcagctgcgtgttgctgaataatgctcttccacagacc 1895
 Db 561 GCCGAGAGCTGGTAGTGTCTTCAGTTCGGTGTTCCTAACATGGCTTCTCCAACGACC 620
 QY 1896 tgttcaacaagagctctctgagtagcagcagctctgagcaacaattcacacagctgctg 1954
 Db 621 TGTTCACAAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACAGACCTGGTG 679

RESULT 4

AL157379 194704 bp DNA HTG 22-MAR-2000
 LOCUS Homo sapiens chromosome 6 clone RPL-62L18, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 28 unordered pieces.
 ACCESSION AL157379
 VERSION AL157379.2 GI:7159486
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 194704)
 AUTHORS Sims,S.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT requests: clonerequest@sanger.ac.uk
 On Mar 6, 2000 this sequence version replaced gi:7009540.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in
 progress and the release of this data is based on the understanding
 that the sequence may change as work continues. The sequence may
 be contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc. Order of segments is not known; 800 n's separate

DEFINITION Macaca fascicularis interphotoreceptor matrix proteoglycan 150 mRNA, partial cds.
ACCESSION AF047491
VERSION AF047491.1 GI:2906229
KEYWORDS crab-eating macaque.
SOURCE Macaca fascicularis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 555)
AUTHORS Kuehn, M.H. and Hageman, G.S.
TITLE Characterization And Complete cDNA Sequence Of IPM 150, A Novel Human Photoreceptor Cell-Associated Chondroitin-Sulfate Proteoglycan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 555)
AUTHORS Kuehn, M.H. and Hageman, G.S.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1998) Ophthalmology and Visual Sciences, University of Iowa, 200 Hawkins Drive, 11190E PFP, Iowa City, IA 52240, USA
FEATURES Location/Qualifiers
source 1..555
ORIGIN /organism="Macaca fascicularis"
DB /db_xref="taxon:9541"
CD /tissue="retina"
note <1..>555
codon_start 1
product "interphotoreceptor matrix proteoglycan 150"
protein_id "AAC03788.1"
db_xref "GI:2906230"
translation "IFPNGVKVCPQESMKQILASQAYVLRVCOEAVWEYRIFLD RLPDGEYDWSFCQEFCLFDICQNSQEHDLQORIKQSFPRKDEVSTE RTGSEYFIVSTVDVSLGPPVPTDPTLLNEILDNALDKPTTRETVELAVS EQRVLESISLNRFKELADQS"
BASE COUNT 162 a 131 c 131 g 131 t
ORIGIN
Query Match 15.7%; Score 511; DB 11; Length 555;
Best Local Similarity 95.5%; Pred. No. 1.5e-105;
Matches 526; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 344 tttttcccaacggggttaagctgtctccacaggaatccatgaacagatttttagacagt 403
DB 4 tttcttccgaacgggttaagctgtctccacaggaatccatgaacagatttttagacagt 63
QY 404 ctccaagcttattagattgagtggtgtcaggaagcagatgggaagcatatcgagtc 463
DB 64 ctccaagcttattagattgagtggtgtcaggaagcagatgggaagcatatcgagtc 123
QY 464 tttctgagtcgcatccctgacacaggggaataatcagagctgggtcagcatctgcagcag 523
DB 124 tttctgagtcgcatccctgacacaggggaataatcagagctgggtcagcatctgcagcag 183
QY 524 gagaccttctgctctttgacattggaaaaacttcagcaattcccccaggagcactggat 583
DB 184 gagaccttctgctctttgacattggaaaaacttcagcaattcccccaggagcactggat 243
QY 584 cttctccagcagagaataaacagagaagtttccctgacagaaaaagatgaataatctgca 643
DB 244 cttctccagcagagaataaacagagaagtttccctgacagaaaaagatgaataatctgca 303
QY 644 gagaacattggagagcctgtgaaaccattgtcattcaacagatgttgcaacgctc 703
DB 304 gagaacattggagagcctgtgaaaccattgtcattcaacagatgttgcaacgctc 363
QY 704 tcaattgggcttcttccctctcactcctgtagacacccctcctcaatgaattctcgataat 763
DB 364 tcaattgggcttcttccctctcactcctgtagacacccctcctcaatgaattctcgataat 423

QY 764 acactcaacgacacacaaagatgctctacacagaaaagagaaagattcgctgtgttgag 823
DB 424 GCACCTAACGACACCAAGATGCTTACAACAGAAAGAAACAGAACTCGCTGTCTGAG 483
QY 824 gagcagagggtgagctcagctctctggttaacacagaaattcaaggcagagctcgt 883
DB 484 GAGCAGAGGGTGGAGCTCAGCATCTCTGTATAAACAGAGAGGTTCAAGCAGAGCTGCT 543
QY 884 gactcccgatc 894
DB 544 GACTCTCAGTC 554
RESULT 6
HSIMPG10 444 bp DNA PRI 28-OCT-1998
DEFINITION Homo sapiens interphotoreceptor matrix gene (IPM150), exon 10.
ACCESSION AF017769
VERSION AF017769.1 GI:3800724
KEYWORDS 10 of 17
SEGMENT human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS Feilbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M., Schmid, M. and Weber, B.H.
TITLE Genomic organization and chromosomal localization of the interphotoreceptor matrix proteoglycan-1 (IPMGI) gene: a candidate for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 444)
AUTHORS Gehrig, A., Feilbor, U., Kelsell, R., Hunt, D.M., Maunee-Hussels, I.E. and Weber, B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150) localized to 6q14.2-q15 in autosomal dominant Stargardt-like macular dystrophy, progressive bifocal choroidretinal atrophy (PBCRA), and North Carolina macular dystrophy (MCDRI)
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 444)
AUTHORS Feilbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am Hubland, Wuerzburg D-97074, Germany
FEATURES Location/Qualifiers
source 1..444
ORIGIN /organism="Homo sapiens"
DB /db_xref="taxon:9606"
CD /chromosome="6"
map "6q14.2-q15"
base 92..339
gene "IPM150"
BASE COUNT 133 a 86 c 101 g 124 t
ORIGIN
Query Match 7.6%; Score 249; DB 11; Length 444;
Best Local Similarity 100.0%; Pred. No. 3.9e-46;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1014 gctcaagctccacagagatgcaacttacggccatctttaagagacacagtcagagaagaa 1073
DB 91 GCTCAAGCTCCACAGAGATGCAACTTACGGCATCTTTTAAGAGACACAGTGCAGAGCAA 150
QY 1074 aaagccctgcaagtgacctctgtctttgattccacaaaattgaagtgagagctc 1133
DB 151 AAAGCCCTGCAAGTGACCTCTCTGCTTTTATTCCAAATAATTGAAGTGAAGAGTCT 210
QY 1134 atcatggaacatgagagagcaagcaacacagaaatctatctcacagctacagacctca 1193
DB 211 ATCATGGNACCATGGAGGAGGACCAAGCAACCAAGCAATCTATCTCACAGCTACAGACCTCA 270

* 28111 28010: gap of 800 bp
* 28911 32475: contig of 3565 bp in length
* 32476 33275: gap of 800 bp
* 33276 38322: contig of 5047 bp in length
* 38323 39122: gap of 800 bp
* 39123 40287: contig of 1165 bp in length
* 40288 41087: gap of 800 bp
* 41088 43973: contig of 2886 bp in length
* 43974 44773: gap of 800 bp
* 44774 60308: contig of 15535 bp in length
* 60309 61108: gap of 800 bp
* 61109 81103: contig of 19935 bp in length
* 81104 81903: gap of 800 bp
* 81904 82931: contig of 1028 bp in length
* 82932 83731: gap of 800 bp
* 83732 91019: contig of 7288 bp in length
* 91020 91819: gap of 800 bp
* 91820 98660: contig of 6841 bp in length
* 98661 99460: gap of 800 bp
* 99461 100575: contig of 1115 bp in length
* 100576 101375: gap of 800 bp
* 101376 114138: contig of 12763 bp in length
* 114139 114938: gap of 800 bp
* 114939 138142: contig of 23204 bp in length
* 138143 138942: gap of 800 bp
* 138943 146423: contig of 7481 bp in length
* 146424 147223: gap of 800 bp
* 147224 154028: contig of 6805 bp in length
* 154029 154828: gap of 800 bp
* 154829 156288: contig of 1460 bp in length
* 156289 157088: gap of 800 bp
* 157089 167098: contig of 10010 bp in length
* 167099 167898: gap of 800 bp
* 167899 170614: contig of 2716 bp in length
* 170615 171414: gap of 800 bp
* 171415 173697: contig of 2283 bp in length
* 173698 174497: gap of 800 bp
* 174498 184556: contig of 10059 bp in length
* 184557 185356: gap of 800 bp
* 185357 187387: contig of 2031 bp in length
* 187388 188187: gap of 800 bp
* 188188 189966: contig of 1779 bp in length
* 189967 190766: gap of 800 bp
* 190767 192406: contig of 1640 bp in length
* 192407 193206: gap of 800 bp
* 193207 194704: contig of 1498 bp in length.

FEATURES

Location/Qualifiers
1..194704
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RPI-62L18"
/clone_lib="RPCI-1"

BASE COUNT 53528 a 32672 c 33519 g 53383 t 21602 others
ORIGIN

Query Match 7.3%; Score 237.4; DB 32; Length 194704;
Best Local Similarity 99.6%; Pred. No. 1.7e-43;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 190 taaagatatctccattaaacataacattctgaacattgaacataaagacataagacataatcccccaag 249
|||
DB 132030 TACAGTATCTCCATTAAACATATACCATCTGAACTTAAAGACATAGACATCCCCCAAG 131971
|||
QY 250 aaatgaacaactgaagtactgaaataatgtacaaatgtcaaatgtcaactgatgacgaatatt 309
|||
DB 131970 AAATGAACAACGTGAAGTACTGAAAATGTACAAAATGTCAACTACTATGACGAATATT 131911
|||
QY 310 cgattggcaagcatcgacacaaaagatccgcatttttcccaacgggggttaagtctg 369
|||
DB 131910 CGATTGGCAAGCATCGACACAAAAGATCCGCATTTTCCCAACGGGGGTAAAGTCTG 131851
|||

QY 370 tccacaggaatccatgaacagatttttagacagcttccaagcttattattagattgagag 428
|||
DB 131850 TCCACAGGAATCCATGAACAGATTTTACAGAGCTTCAAGCTTATTATAGATTGAGAG 131792
|||

RESULT 9
HSIMPG14 448 bp DNA PRI 28-OCT-1998
LOCUS Homo sapiens interphotoreceptor matrix gene (IPM150), exon 14.
DEFINITION AF017773
ACCESSION AF017773.1 GI:3800728
VERSION AF017773.1
KEYWORDS i4 of 17
SEGMENT human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 448)
AUTHORS Felbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
Schmid, M. and Weber, B.H.
TITLE Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPc1) gene: a candidate
for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 448)
AUTHORS Gehrig, A., Felbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussells, I.E.
and Weber, B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choriorretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRI)
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 448)
FELBOR, U., KUEHN, M., HAGEMAN, G.S. and WEBER, B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES
Source 1..448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
111..330
/gene="IPM150"

exon
BASE COUNT 146 a 85 c 89 g 128 t
ORIGIN

Query Match 6.9%; Score 226; DB 11; Length 448;
Best Local Similarity 90.6%; Pred. No. 6.5e-41;
Matches 241; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1930 ggagcaacaattcacacagctggttccattatctacgattcccaattcttacaggatttaa 1989
|||
DB 89 GAAACCATCTTTATTTTGAGCTGGTTCATATCTAGATCCCAATCTTACAGGATTTAA 148
|||
QY 1990 gcaactgaaataacttaacttcagaacggaggagtgtgattgtaatacgaataaagt 2049
|||
DB 149 GCAACTTGAATATCTTAACCTTCAGAAACGGGAGTGTGATTGTGAATACGAAATCAAGTT 208
|||
QY 2050 tgctaagtctgtgcgtataacacctcacaaggctgtgcacggggtctctggaggatttcg 2109
|||
DB 209 TGCFAAGTCAGTGGCGTATAACCTCACCAGGCTGTGCACGGGTCTTTGGAGGATTTTCG 268
|||
QY 2110 ttctgctgaccccaacaactccattctggaataagacagctactctccaacattgaacc 2169
|||
DB 269 TTCTGCTGACGCCCAACAACATCCATCTCGAAATAGACACTACTCTCTCAACATTGAACC 328
|||
QY 2170 agctgatcaagcagatccctgcaagt 2195
|||
DB 329 AGGTAAAGAAATCCCAACCCAGAAAGT 354
|||


```

RESULT 10
AF157624
LOCUS      AF157624      3989 bp      mRNA      PRI      06-MAR-2000
DEFINITION Homo sapiens Spacrcan mRNA, complete cds.
ACCESSION  AF157624
VERSION     AF157624.1  GI:6467402
KEYWORDS   .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3989)
AUTHORS     Acharya,S., Foletta,V.C., Lee,J.W., Rayborn,M.E., Rodriguez,I.R.,
            Young,W.S. 3rd and Hollyfield,J.G.
TITLE       SPACRCAN, a novel human interphotoreceptor matrix
            hyaluronan-binding proteoglycan synthesized by photoreceptors and
            pinealocytes
JOURNAL     J. Biol. Chem. 275 (10), 6945-6955 (2000)
MEDLINE     20167166
REFERENCE   2 (bases 1 to 3989)
AUTHORS     Foletta,V.C. and Young,W.S.
TITLE       Direct Submission
JOURNAL     Submitted (09-JUN-1999) SNGE, NIMH, 36 Convent Dr, MSC 4068,
            Bethesda, MD 20892-4068, USA
FEATURES    Location/Qualifiers
            1..3989
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="3"
             /map="3q11.2"
            15..3740
             /note="hyaluronan-binding proteoglycan; interphotoreceptor
             matrix component"
            /codon_start=1
             /product="Spacrcan"
             /protein_id="AAF13154.1"
             /db_xref="GI:6467403"
             /translation="MIMFLPGKISGLILFVLEGDPPLTAQTYLSIERIQPKSA
            NFKYFRVCQAEAWFAKFTFDRLPGREYHYMNLCEDGVTSIFEMGNFSEIVE
            HSFLMKKLYAKTSVSSLSPPVGDVTLGDTLTSVPHPVDVAHQSESLER
            PESISNEIENVEAKPQAGEIAEFISHLGKQYREELQDSFVHGHLEFSE
            VENAFTGLPGYKELRVLEFSPKENDSDVVVAVTNGEAISSNTWDLISLHNKVE
            NHGVLDDRPVTVYITISNFDYIAETLQONFLGNSLNPDPQLSLINVRGLRHQ
            TEDLVWNTQSSSLQATPSSILNDFQAAPNSADESITSSIPPLDFSSGPPSATRELW
            SESPLGLVTHKLAFTPSKMLSSPEVLEYSLSLHSTVPAVLQTLGPPVASERTSG
            SHVEDGLANVEEEDSLSDLPSSFTQVPKETIPSMEDSDVSLTSSPLTSSIP
            FGLDLSKVKDQLKVPFLPDASMEKELIFDGLGSGGOKVLDIOWSKHKYVHDDRSIHPEEPEL
            AEPLSKPLEDSDSLPAELEDKLLVLDKNDSDIOLSKHKYVHDDRSIHPEEPEL
            SGPAVPIADTAASASLTLPKHISEVPGVDVDYVTRAPLILTSVAISASDKDQAD
            ALRDEMEQITESNYEMFDESEVSNVXPDMLTILPESEVWTRTSSLEKSLKDL
            AITPQSDARLWLSVTQTKLPPTTISTILLEDEVINGVQDISLELDRTGTQYEPQV
            EQNGKGVSEVMSVSTSEVMSVAVPTEGGDLSYTTQSGALVFFSLRYTNMFSE
            DLFNKLSLEKALQORLELIVLPYLOSNTLGFQNLILNFRNGSIVTVNSRKFANSVP
            PUNNAVYMLIEDCTTAYTNMLAIDKSLDVSDEANPCKQACNEFSECLVNPW
            SGENKRCFPGLTVERPCQSLDLPDFCNQDKDIMPCHGAICRCRVGENWYIR
            GKHOEEFVPEIIGITIASVGLLVIFSAIYFFITLQAHDRSERESPFSGSRQ
            PDSLSIENAVKYNPVYSHRAGCEYEGYPQHPFYSASGSDVIGGLSREIRQMYE
            SSLSREIQRMVLELYANDPFAFVRQQVEEV"
BASE COUNT 1146 a 829 c 924 g 1090 t
ORIGIN

Query Match      6.2%; Score 203.4; DB 50; Length 3989;
Best Local Similarity 63.7%; Pred. No. 8.6e-36;
Matches 309; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 1818 tgaccattgcccccaaggccagagctgttagtcttcttcagctgtgcgttgctaaaca 1877
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2683 TGAGTTATACCCAGACTTCAGGAGCTTTGGTGGTTTCTTCAGCTCCCGAGTGACTAACA 2742

```



```
Db 3110 CTAATGTCAACAACCCATATATATCTGTGAAGACTTTTGTACCAACCGCTACCAAA 3169
Qy 2127 aactccatctggaatagacagactactctcaacatgaacacagctgatacagagatc 2186
Db 3170 CCATGAACCTGGATATGATAAGTACTCCCTGGATGTGGAATCAGGTGATGACGCCAAC 3229
Qy 2187 cctgcaagtctcctgcgcggaatttgcgaatgtgtaagaacagcagcagactgag 2246
Db 3230 CTTGCAAGTTTCAGCCCTGTATGATTTCTGATGTTTGGTAATCCATCCATGGAGTGGAG 3289
Qy 2247 aagcggagctgcgtgcaaacagcagatgacacgacggcgagcctggagcgtc 2300
Db 3290 AAGCAAAATGCAAAATGCCACCCCTGGTACCTGAGTGTGATGAATGCCTTGTGTC 3343

RESULT 13
HSIMPG15
LOCUS Homo sapiens 477 bp DNA PRI 28-OCT-1998
DEFINITION Homo sapiens interphotoreceptor matrix gene (IPM150), exon 15.
ACCESSION AF017774
VERSION AF017774.1 GI:3800729
KEYWORDS
SEGMENT
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS Felbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
Schmid, M. and Weber, B.H.
TITLE Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPG1) gene: a candidate
for 6q14.2-q15 retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 477)
AUTHORS Gehrig, A., Felbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussells, I.E.
and Weber, B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choroidretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRI)
Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 477)
AUTHORS Felbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humanogenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES
source
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
71..269
/gene="IPM150"
BASE COUNT 127 a 119 c 117 g 114 t
ORIGIN

Query Match 5.8%; Score 189.2; DB 11; Length 477;
Best Local Similarity 93.7%; Pred. No. 1.4e-32;
Matches 208; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

Qy 2151 actctcaacattgaaccagctgacagcagatccctgcaagtcctggcctgcgcg 2210
Db 50 AATCTGCCCATTTCTTACAGCTGATCAACAGAGATCCCTGCAAGTTCCTGGCCCTCGCGCG 109
Qy 2211 aatttgcacattgttaagaacagcagcagcagcagcagcagcagcagcagcagcagcag 2270
Db 110 AATTGGCCCAATGTGTAAAGACGACGAGCTGAGGAGCGGAGTGTGCTGCGTCAAAACGAG 169
```

```
Qy 2271 gatatgacagccaggagcctgcggagcctgcggagcctgcggagcctgcgtg- 2329
Db 170 GATATGACAGCCAGGGAGCGCTGGACGGTCTGGAACAGGCCCTCTGGCCCTGGCACAA 229
Qy 2330 aggaatgcaggtcctccagggaaggagcctccatgcggtt 2371
Db 230 AGGAATGCGAGGTCTCCAGGGAAGGAGGAGCTCCATGCAGGT 271

RESULT 14
HSIMPG01
LOCUS Homo sapiens 310 bp DNA PRI 28-OCT-1998
DEFINITION Homo sapiens interphotoreceptor matrix gene (IPM150), exon 1.
ACCESSION AF017760
VERSION AF017760.1 GI:3800715
KEYWORDS
SEGMENT
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 310)
AUTHORS Felbor, U., Gehrig, A., Sauer, C.G., Marquardt, A., Kohler, M.,
Schmid, M. and Weber, B.H.
TITLE Genomic organization and chromosomal localization of the
interphotoreceptor matrix proteoglycan-1 (IMPG1) gene: a candidate
for 6q-linked retinopathies
JOURNAL Cytogenet. Cell Genet. 81 (1), 12-17 (1998)
MEDLINE 98358139
REFERENCE 2 (bases 1 to 310)
AUTHORS Gehrig, A., Felbor, U., Kelsell, R., Hunt, D.M., Maumenee-Hussells, I.E.
and Weber, B.H.F.
TITLE Assessment of a novel interphotoreceptor matrix gene (IPM150)
localized to 6q14.2-q15 in autosomal dominant Stargardt-like
macular dystrophy, progressive bifocal choroidretinal atrophy
(PBCRA), and North Carolina macular dystrophy (MCDRI)
Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 310)
AUTHORS Felbor, U., Kuehn, M., Hageman, G.S. and Weber, B.H.F.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1997) Humanogenetik, Universitaet Wuerzburg, Am
Hubland, Wuerzburg D-97074, Germany
FEATURES
source
1..310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q14.2-q15"
<1..223
/gene="IPM150"
<1..156
/gene="IPM150"
BASE COUNT 108 a 47 c 45 g 110 t
ORIGIN

Query Match 5.3%; Score 174.4; DB 11; Length 310;
Best Local Similarity 84.6%; Pred. No. 3.3e-29;
Matches 219; Conservative 0; Mismatches 37; Indels 3; Gaps 2;

Qy 1 taacaacaagaaggttatccctcaatcatctggtatcaataataattttt--cctttt 58
Db 27 TAAACCAAGAGGTTATCCCTCAATCATCTGCTATCAATATATATATATTTTTCACATTTC 86
Qy 59 tgttacttttttaagattgaggttgcgtggttgcgtgattgtttatcagaattacc-atgcac 117
Db 87 TGTACTTTTAAATGAGATTGAGTTGAGTTGTTCTGTGATTGTTATCAGAAATACCAATGCAC 146
Qy 118 aaagccagaatgtatttggaactagaagcagcagcagcagcagcagcagcagcagcagcag 177
Db 147 AAAAGCCAGAATGATTTGGAAACTAGAGAGCTATTTTGTGGATTTCCTTCTCCA 206
Qy 178 agttcaaggaactaaagatatctccatcataataaccattctgaaacataagacataga 237
```

Db 207 AGTTCAGGACCAAGGTAAGTTACTTAAATGCTTTTAAATGCTTATCTATAA 266
 QY 238 caatcccccaagaatgaa 256
 Db 267 ATCTACCGATAGAAAGTGA 285

RESULT 15

HSIMPG03	HSIMPG03	422 bp	DNA	PRI	28-OCT-1998
LOCUS	Homo sapiens interphotoreceptor matrix gene (IPM150), exon 3.				
DEFINITION	AF017762				
ACCESSION	AF017762.1	GI:3800717			
VERSION					
KEYWORDS	3 of 17				
SEGMENT	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 422)				
AUTHORS	Felbor,U., Gehrig,A., Sauer,C.G., Marquardt,A., Kohler,M., Schmid,M. and Weber,B.H.				
TITLE	Genomic organization and chromosomal localization of the interphotoreceptor matrix proteoglycan-1 (IMPGL) gene: a candidate for 6q-linked retinopathies				
JOURNAL	Cytogenet. Cell Genet. 81 (1), 12-17 (1998)				
MEDLINE	98358139				
REFERENCE	2 -(bases 1 to 422)				
AUTHORS	Gehrig,A., Felbor,U., Kelsell,R., Hunt,D.M., Maumenee-Hussels,I.E. and Weber,B.H.F.				
TITLE	Assessment of a novel interphotoreceptor matrix gene (IPM150) localized to 6q14.2-q15 in autosomal dominant Stargardt-like macular dystrophy, progressive bifocal choriorretinal atrophy (PBCRA), and North Carolina macular dystrophy (MCDRI) Unpublished				
JOURNAL	3 (bases 1 to 422)				
REFERENCE	Felbor,U., Kuehn,M., Hageman,G.S. and Weber,B.H.F.				
AUTHORS	Direct Submission				
TITLE	Submitted (09-AUG-1997) Humangenetik, Universitaet Wuerzburg, Am Hubland, Wuerzburg D-97074, Germany				
JOURNAL	Location/Qualifiers				
FEATURES	source				
	1..422				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="6"				
	/map="6q14.2-q15"				
	112..278				
	/gene="IPM150"				
exon					
BASE COUNT	113 a 90 c 103 g 116 t				
ORIGIN					
Query Match	5.2%; Score 169.8; DB 11; Length 422;				
Best Local Similarity	98.8%; Fred. No. 3.6e-28;				
Matches	171; Conservative 0; Mismatches 2; Indels 0; Gaps 0				
QY	423	tgagagttgtcaggaacgacgatgggaagcatatcggtatcttctggatcgatccctg	482		
Db	106	TTACAGTGTGCAGGAGCAGATATGGGAGCATATCGGATCTTTCTGGATCGATCCCTG	165		
QY	483	acacaggggaatatcaggactgggtcagcatctgcacgacgagacacctctgcctctttg	542		
Db	166	ACACAGGGGATATCAGGACTGGGTTCAGCATCTGCACGAGGAGACCTTCTGCCTCTTTG	225		
QY	543	acattggaaaaacttcagcaattccccaggagcacctggatctttctccacgag	595		
Db	226	ACATTGGAAAAAACCTTCAGCAATTCCTCCAGGACACCTGATCTTCTCCACGAG	278		
